DOCUMENT RESUME

ED 360 349 TM 020 229

TITLE	Consultation on Evaluation of Youth Development Programs. Report on the Meeting (January 15, 1992).
INSTITUTION	Carnegie Council on Adolescent Development, Washington, DC.
PUB DATE	15 Jan 92
NOTE	77p.
PUB TYPE	Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Adolescents; *Community Programs; Conferences; *Disadvantaged Youth; Elementary Secondary Education; Evaluation Methods; *High Risk Students; Models; Philanthropic Foundations; *Program Evaluation
IDENTIFIERS	Carnegie Council on Adolescent Development; *Youth Development Program; Youth Opportunity Program

ABSTRACT

In 1990, the Carnegie Council on Adolescent Development convened a Task Force to guide the work on a new Project on Youth Development and Community Programs. The major goals of this project are to expand the scope and availability of developmentally appropriate, community-based services for young adolescents, particularly those in high-risk environments. Two of the specific mandates of the project were related to the evaluation of youth programs. On January 15, 1991, a one-day consultation was held to assess the current challenges and successes of youth organizations as they work to evaluate their programs and to make recommendations for strengthening program evaluation efforts. This document contains: (1) a summary of the meeting; (2) a roster of the approximately 20 participants (Appendix A); (3) a summary of their written answers to questions (Appendix B); (4) a summary of the state of program evaluation within 19 selected national youth organizations (Appendix C); (5) summaries of 3 selected articles included in the briefing report preparing participants for the meeting (Appendix D); and (6) a bibliography of 22 sources on evaluation of youth development programs (Appendix E). (SLD)

****
*   Reproductions supplied by EDRS are the best that can be made   *
*   from the original document.   *
****

TM

CARNEGIE COUNCIL ON
ADOLESCENT DEVELOPMENT

A PROGRAM OF CARNEGIE CORPORATION OF NEW YORK

ED 360 349

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Ruby Takanishi

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

REPORT ON THE CONSULTATION ON
EVALUATION OF YOUTH DEVELOPMENT PROGRAMS

January 15, 1992

Carnegie Corporation of New York

Carnegie Council on Adolescent Development

Task Force on Youth Development and Community Programs

BEST COPY AVAILABLE

TM 020229

2400 N St., N.W., 6th floor
Washington, D.C. 20037-1153
(202) 429-7979
FAX (202) 775-0134

David A. Hamburg, Chairman
H. Keith H. Brodie
Michael I. Cohen
Alonzo A. Crim
Michael S. Dukakis
William H. Gray III
Beatrix A. Hamburg
David E. Hayes-Bautista
Fred M. Hechinger
David Hornbeck
Daniel K. Inouye
James M. Jeffords
Richard Jessor
Helene L. Kaplan
Nancy L. Kassebaum
Thomas H. Kean
Ted Koppel
Hernan LaFontaine
Eleanor E. Maccoby
Ray Marshall
Julius B. Richmond
Frederick C. Robbins
Kenneth B. Smith
Wilma Tisch
P. Roy Vagelos
William J. Wilson

INTRODUCTION

The Carnegie Council on Adolescent Development, an operating program of the Carnegie Corporation of New York, was created in 1986 to advocate for the higher placement of adolescent issues on the national agenda. In 1990 the Council convened a 27-member Task Force to guide the work on a new Project on Youth Development and Community Programs. The two major goals of the project are to expand the scope and availability of developmentally appropriate, community-based services for young adolescent (ages 10-15), particularly those living in high-risk environments; and to enhance public understanding and support of effective services for America's youth.

The Project's mandate was to address twelve specific objectives, two of which were related to program evaluation:

- To assess proven and promising approaches to serving disadvantaged youth through community programs; and
- To stimulate positive change within youth-serving organizations themselves and within other institutions that promote youth development as a portion of their work.

As the Task Force began its work, it became increasingly clear that although there had been some progress made over the last 10 years, not enough was known about the actual impact of youth development activities on the lives of young people themselves. Given this finding, the Task Force decided to convene a meeting to be attended by representatives from three groups concerned about program evaluation in youth organizations: evaluation experts, youth development professionals and funders. In November of 1991, approximately twenty key professionals were invited to submit written comments and to attend a one-day meeting (consultation) designed to:

- document and assess the current challenges and successes experienced by youth organizations as they work to evaluate their program efforts; and
- to make and prioritize recommendations for strengthening program evaluation efforts within youth organizations.

The meeting was held on January 15, 1992 and was attended by 21 participants and two observers.

This document contains: (1) a summary of the meeting itself; (2) a roster of consultation participants; (3) a summary of participants' written comments, (4) a summary of the state of program evalution within selected national youth organizations; (5) summaries of selected articles included in the briefing report that prepared participants for the meeting; and (6) a bibliography on evaluation of youth development programs.

CARNEGIE COUNCIL ON ADOLESCENT DEVELOPMENT
CONSULTATION OF EVALUATION OF YOUTH DEVELOPMENT PROGRAMS
MEETING SUMMARY

The Consultation on Evaluation of Youth Development Programs took place on January 15, 1992 at the Carnegie Corporation of New York. Judith Torney-Purta, a member of the Task Force on Youth Development and Community Programs, served as chair.

The format of the Consultation consisted of five segments. In the first segment, participants introduced themselves and articulated what they saw as the single most important issue to address in order to strengthen program evaluation in the field of youth development. In segment two panelists representing three national youth organizations described their experience carrying out a major outcome evaluation. For each of the three featured case studies, an administrator from the youth organization was paired with the evaluator for the study, and the two offered their individual perspectives on the successes and challenges encountered during the project. In segment three the entire group of participants responded to a set of questions proposed by the chair. The fourth segment featured another panel composed of representatives from three different funding agencies who presented their varying perspectives on program evaluation. In the fifth and final segment, participants worked in small groups to identify their top three recommendations for strengthening program evaluation efforts within youth organizations.

SEGMENT ONE: MOST IMPORTANT ISSUES TO ADDRESS

Although many remarked that it was a great challenge to undertake, participants identified the following issues as the single most important ones to address in order to strengthen program evaluation in the field of youth development:

1. Many of the nation's oldest and largest youth organizations, such as Boy Scouts of America and Girl Scouts of the U.S.A., have not taken the time or allocated the resources to focus on outcome evaluation. What is the long term impact of participation in these and other youth development programs?

2. Lacking adequate documentation youth organizations make educated guesses and sometimes even inflated claims about the impact of their programs. As a first step, efforts must be made to identify the outcomes that would be the expected result of successful participation in youth development activities. It is time to clarify the areas in which youth development programming does have an impact and under what conditions that impact is greatest. When identifying outcomes, it is important to keep in mind the diversity of program goals, objectives, activities, clientele and so on.

Success may be viewed very differently from agency to agency and from community to community. What range of specific outcomes is realistic for practitioners, researchers and funders?

3. Not only have youth development organizations fallen into the trap of making inflated claims, but the inflated claims tend to focus on what some believe are the "wrong" outcomes. Due to the categorical nature of funding streams, many agencies develop programs aimed at preventing or reducing problem behaviors such as violence, drug use or adolescent pregnancy. Although some of these programs have proven to be successful in preventing or reducing problem behaviors, it is time for the field to identify a consistent list of positive outcomes to be achieved by youth organizations. What are the positive, rather than preventative, outcomes that should result from participation in youth development activities? How can achievement of these positive outcomes be linked to the reduction of negative outcomes so youth organizations can demonstrate that they do have an impact (both directly and indirectly) on adolescent problems?

4. Evaluation is more a state of mind than a set of tools. If program staff people see evaluation as an external, threatening or hostile activity, they are more likely to resist it. Once youthworkers, volunteer or paid, have a clearer understanding of the outcomes their programs are designed to achieve, then if evaluation is not automatic it is at least welcome. What staff development efforts must be put in place for youthworkers so that evaluation follows naturally and seems normal? How do we help staff appreciate that evaluation results will recognize and support the work they're doing on a day-to-day basis?

5. Many proposals to funders from youth organizations have weak evaluation designs. Often evaluation is an afterthought rather than an integral component of good program design. What can be done to enable practitioners to pay attention to evaluation from the very beginning as they design program goals and objectives? How can we empower local indigenous youth organizations to do their own ongoing assessments of their programs?

6. In the foundation world there are sometimes conflicts among program officers, boards of directors and grant recipients about the most desirable approach to program evaluation. What can be done to establish some common goals and evaluation criteria for all stakeholders?

7. Foundations that are interested in research spend a lot of time trying to work with youth organizations and evaluators to effect a "marriage," and yet the "divorce rate" is high.

What guidelines can be offered to help create more mutually-rewarding relationships between youth organizations and outside evaluators? How should evaluators structure their research and reports to be of maximum use in program development efforts?

8. Funders have increasingly come to insist that program evaluation be both theoretically and methodologically sound. It takes time and resources for evaluators and practitioners to work together as partners in the creation of such evaluation designs. It takes time for an evaluator to come into an agency, gain insight into the organizational culture, hear from practitioners what impact they think they are having, identify appropriate outcomes against which a program ought to be measured, and come up with evaluation measures that would be a fair test. How willing are funders to provide the financial support and have the patience that will allow evaluators and program staff to do the preliminary work to make evaluations both programmatically and methodologically sound?

9. It is essential to bridge the gap between evaluators and practitioners. Often when program staff sit around the same table with evaluators, the practitioners tend to feel intimidated. When the Education Development Center (EDC) reviewed the state of the art of evaluation of violence prevention programs, practitioners expressed several concerns. First, they indicated that their premises about outcomes were frequently discounted by evaluators who were quick to say, for example, "the research doesn't bear that out." Second, program staff commented on the cultural and educational schism between themselves and their "high-risk" clients, on the one hand, and the outside evaluators who were typically highly educated European Americans on the other. What steps can be taken to bridge the cultural, professional and philosophical gaps between evaluators and practitioners?

10. Whether the issue is professional development of staff or program evaluation, fragmentation is a big problem. Youth organizations rarely have opportunities to share their challenges and successes related to evaluation. Funders, such as the federal government, have much to say about the need for coordination among grant recipients but don't actually do very much to promote such coordination. What steps can be taken by funders, national and local youth organizations and evaluators to encourage more collaboration on this issue?

11. Too often, the outcomes generated by evaluations of youth development programs are not embraced by both the scientific and programmatic communities. If a program is shown to be

effective, some scientists rush to criticize the methodology--saying for example, the control group wasn't comparable, the measures weren't objective, or the evaluators were insiders who were biased toward positive outcomes. If the evaluation results are negative, then the scientific community might tend to believe that the particular program did not have an impact, while the programmatic community might tend to believe that the evaluation did not capture the actual results. Until youth development organizations demonstrate the success of their programs through rigorous evaluation methods and gain a legitimate voice not only on the programmatic side, but also on the research side, it will be difficult to attract the funding, the national support, and the attention that these programs richly deserve. Which evaluation methods will be sufficiently flexible to address the realities of youth development programs, but also uncompromisingly rigorous so that the outcomes, the results generated by those methods, are accepted and respected by both the scientific and programmatic communities? How can youth development organizations use evaluation to elevate the understanding and communicate the value of their organizations to policymakers and to national and local funding sources?

12. There are different levels of evaluation that should be undertaken in youth organizations. It is probably reasonable to expect a typical organization to ask and answer its own questions about the quantity and quality of services and to document the process of what goes on in their agencies in order to create feedback to improve programs. But when it comes to questions about effectiveness in changing the lives of individual youth or questions about the transferability of effective programs from site to site, it might be unreasonable to expect such evaluations to be conducted on a routine basis. These kinds of evaluations are very expensive to do well and they require the skill and experience of seasoned professionals. Which kinds of evaluation questions can one reasonably expect a youth organization to answer by itself with limited financial and human resources? How often and under what conditions should youth development organizations undertake large scale, high quality evaluations, looking for conditions of effectiveness and transfer?

13. Too often in discussions of program evaluation the focus is solely on the impact of program activities such as those aimed at skill-building. Youth development organizations don't just build skills. They also create a solid environment in which young people can be accepted as they are, make mistakes and learn from them, participate in making decisions and carrying them out, take reasonable risks, and most importantly, have relationships with caring

adults. New evaluation efforts must investigate the impact of particular environments and staff-youth relationships. What is it that staff do for and with young people? What are the important qualities of those relationships and how can we document them? How can these findings be integrated into staff training and staff development?

14. There is a difference between the impact of the core activities of a youth organization and the impact of a particular categorical program. It is a mistake to use the words program and organization interchangeably in discussions of evaluation. There are youth development organizations that have as their mission the promotion of well-being of young people. These organizations also implement youth development programs. But many other types of organizations, such as schools, also implement youth development programs. What is the added value of having youth development programs implemented in youth development organizations? A good example is the recent Girls Incorporated evaluation of their adolescent pregnancy prevention program. Theirs was a solid pregnancy prevention program that demonstrated good results. One might wonder whether the program was better able to demonstrate results because it was implemented within a youth development organization. If a school chose that program and replicated it faithfully, it might not be able to demonstrate the same results. What is the impact of the total experience of a youth development organization, as compared with participation in concrete time-limited youth development programs?

15. Most evaluation designs involve asking youth to complete questionnaires or to participate in interviews. It is challenging to get the full participation of youth and to track them as they go through a particular program. What is the best way to gain the cooperation of youth in the evaluation process? What are the best ways to expand the repertoire of information-gathering mechanisms?

16. Most of the evaluations funded by the Ford Foundation are carried out within the context of fairly large scale research demonstration projects that tend to focus on student outcomes such as earnings and employment. These demonstrations yield few insights beyond the impact findings. For instance, an evaluation reveals that the experimental group's earnings have increased more than the control group's, but does not allow attributions as to why that happened. Ford is realizing that there is a need for better theories about why and how humans change. How does the field integrate the data generated from demonstration projects with insights from some of the behavioral sciences (adolescent development theory, learning theory, behavioral

theory, environmental impact research) to develop better conceived interventions?

17. Systematic qualitative studies provide three important insights to the field of youth development. First, these studies are the best way to understand what the inputs are, what the program consists of, how it is actually carried out, what is really done. Second, they provide information about the impact of the program on specific youth. Third, qualitative studies offer powerful and accurate answers to questions about why a particular program does or doesn't accomplish its objectives. What is the best way to convince practitioners, funders and evaluators of the value and potential rigor of qualitative evaluation methods?

18. Youth tend to float in and out of different organizations and programs within the youth development sector. That is a very important feature of youth development, a young person's ability to move voluntarily in and out of these organizations. And yet, most evaluation designs view this movement of youth as a contaminating factor. When evaluations focus exclusively on one program or one organization and its impact on young people, we remove one of the strengths of the youth development sector. What innovative evaluation designs and methods could incorporate the movement of youth from organization to organization as a strength and still deliver credible results?

19. There is a need for a mechanism or process that gives practitioners, funders and evaluators an opportunity to hear about and critique evaluations taking place in youth development organizations. Such open forums would promote the discussion of issues such as "control" and "treatment" groups, random assignment, validity of findings and appropriate data-gathering techniques.

SEGMENT TWO: EVALUATION PANEL DISCUSSION

The next segment offered an opportunity for participants to gain an insider's view of three major youth development evaluation studies. The panel discussion featured:

The Big Brothers/Big Sisters of America (BB/BSA) Evaluation

Dagmar McGill, Deputy National Executive Director, BB/BSA
Alvia Branch, Vice President and Director of Qualitative Research, Public/Private Ventures

SMART Moves: Boys and Girls Clubs of America (B&GCA)
Roxanne Spillett, Assistant National Director of Program Services, B&GCA
Steven Schinke, Professor, Columbia University School of Social Work

Friendly PEERsuasion, Girls Incorporated
Heather Johnston Nicholson, Director, National Resource Center, Girls Incorporated
Marcia Chaiken, Director of Research, LINC

The pairs of panelists discussing each of the evaluation case studies responded to the following set of questions:

a. What were the strengths and limitations of your setting for conducting outcome evaluation?

b. How well did the evaluation design and approach match the underlying assumptions and culture of the youth organization?

c. What outcomes did this evaluation seek to measure and how were these outcomes chosen?

d. How much did the evaluation meet your needs and expectations?

e. What lessons did you learn from this evaluation process? (In your answer, include ways evaluation results have been utilized for program planning and improvement.)

Highlights from the Big Brothers/Big Sisters of America (BB/BSA) Study

BB/BSA is a national organization that provides opportunities for youth to form constructive relationships with supportive adults. Men and women from the community are screened and matched with boys and girls who have been identified as needing additional adult guidance. The organization has 500 affiliates across the country -- in rural, suburban and metropolitan communities, serving approximately 100,000 boys and girls in equal numbers. Public/Private Ventures (P/PV) is a national, not-for-profit organization that seeks ways to improve policy and practice in helping the nation's disadvantaged youth become productively employed and self-sufficient. P/PV joined forces with BB/BSA in 1990 to evaluate the BB/BSA program model. Because it focuses on a relationship rather than a specific program or subject matter, BB/BSA has been very interested in learning about the process of youth development in this model. For P/PV the evaluation of the BB/BSA program is a part of their much larger effort to create a knowledge base related to the effectiveness, cost, operational lessons and best practices associated with providing adult support for at-risk youth. P/PV's research is funded by grants from the Lilly Endowment and an anonymous donor.

Evaluation Design:

The evaluation of BB/BSA involves four separate but interrelated studies:

- a study of the effects of the relationship with a Big Brother or Big Sister (referred to as matches) on the lives of the youth participants;
- a study of volunteer recruitment and screening;
- a study of the administrative and operational practices that comprise the BB/BSA program model, and;
- a qualitative examination of the interactions that take place between the adults and the youth with whom they are paired.

The evaluation, which began in 1990, will be implemented over a four-year period in 15 sites around the country. All of the matches being studied will be completely new to BB/BSA affiliates and will not interviewed until they have had at least two months to form a relationship.

Strengths and Limitations of Setting:

BB/BSA was an appropriate setting for a multi-year, multi-site impact evaluation because:

- it has been in existence for 85 years, has 500 affiliates and serves large numbers of youth.
- BB/BSA operates with goals and objectives -- it knows what it wants to accomplish.
- The organization is familiar with data collection, understands the need for completeness, accuracy, and reliability and, most importantly, doesn't view evaluation as a threat.
- the BB/BSA intervention has major policy implications that should be subjected to a rigorous evaluation.
- BB/BSA maintains a large waiting list of youth, which made it easier to randomly assign a group of interested and eligible youth to a treatment group that would get a Big Brother or a Big Sister, and the remainder to a "wait-list" control group that would not, at least for the period of the study.

Specific Outcome Measures:

Outcome measures were difficult to identify because of the diversity in BB/BSA affiliates across the country. It seemed important not to tailor the outcome measures to the different sets of situations likely to be encountered. Instead, P/PV established a small set of measures to apply to all individual participants in the study. These measures include: 1) self-concept, 2) relationships with parents and peers, 3) self-reported academic performance, 4) reduction of certain anti-social behaviors, and 5) the nature of the relationship maintained between the young person and his/her Big Brother or Big Sister. P/PV went through a long term process of adapting existing measures that are being used in other studies to this particular population. The process consisted of pretesting, extensive consultation with the authors of these measures, and the final tailoring of measures.

Match between Evaluation Design and Underlying Assumptions of the Organization:

P/PV staff spent considerable time getting to know the culture, programmatic methods, and needs of BB/BSA. They attended the BB/BSA national conferences in 1989 and 1990 and participated in many other sessions to discuss issues of importance to BB/BSA such as gender and racial matching, the age of the volunteer, the frequency of contact and tenure of the match, and how all of these factors might impact the child. All of these efforts helped to insure that the questions to be addressed in the evaluation were of interest not only to P/PV and its funders, but also to the BB/BSA national and local staff.

In order to select 15 sites for the project, P/PV carried out a "reconnaissance period," visiting affiliates, interviewing the professional staff, and determining existing levels of interest, support, and data. The goal was to select sites that varied in terms of size, community setting, ethnic makeup of youth served and also to guarantee large enough numbers to make the research effective.

While there was a high level of interest on the part of the agencies to participate, it became clear that the most controversial issue was the utilization of control groups. The idea of not providing service to a group of youngsters who became known to the agency was so hard to justify that one of the metropolitan agencies dropped out of the study at the last minute.

Fulfillment of Expectations:

From the perspectives of both BB/BSA and P/PV, the evaluation has begun successfully. The two organizations have taken the time to understand each other's goals and expectations and to work collaboratively to design methods for gathering the data that they need. Dagmar McGill of BB/BSA expressed complete satisfaction with the project and the relationship with P/PV. Since P/PV had already made a commitment to research the impact of adult-youth mentoring relationships, BB/BSA has turned out to be an excellent organization to provide access to programmatic settings for conducting their research.

Lessons Learned:

The time taken to create a collaborative relationship between P/PV and BB/BSA was time well spent in the sense of contributing to validity and sensitivity in the design.

When asked about other factors in the life circumstances of youth that may have an effect on the outcomes of their relationships with Big Brothers or Big Sisters, Alvia Branch indicated that evaluators will have access to information contained in the case files but will be unable to know with certainty about some of the social context issues or influences affecting the lives of Little Brothers and Little Sisters.

Highlights of the Boys and Girls Clubs of America (B&GCA) SMART Moves Evaluation

B&GCA is a national organization comprised of 1260 affiliated Clubs across the country that serve 1.7 million boys and girls. The Clubs are facility-based -- some are store front operations, others are located in community centers, still others are in public housing projects. Each Club employs at least one staff person who is paid and receives training either locally or from the national organization. The Clubs maintain an open-door policy allowing youth the freedom to come and go, voluntarily. Most come on a fairly regular basis. When children walk through the door of a Boys and Girls Club, they find a variety of activities to choose from and caring staff to facilitate the learning process.

In 1987 the B&GCA national board developed a long range plan to identify and serve more of the young people at greatest risk in this country. The goal was to increase the numbers of youth being served from 1.2 million to 2 million within a specified time period. The board decided on a strategy of establishing new Clubs in public housing complexes around the country. This meant overcoming a host of challenges related to establishing and maintaining operations in the midst of poverty-stricken and often violent communities.

Given this new plan to reach children in public housing, B&GCA began a study funded by the Office for Substance Abuse Prevention (OSAP) to test the following two hypotheses: (1) If a Boys and Girls Club is established in a public housing complex, the Club in and of itself will have some impact on the prevention and/or reduction of substance use. (2) If the SMART Moves drug and alcohol prevention program is implemented in a Boys and Girls Club that is located in public housing, the preventive effect will be enhanced.

Evaluation Design:

In 1987, B&GCA initiated a comparative study that evaluated the effects of Boys and Girls Clubs on children and adolescents who live in public housing and on the overall quality of life in public housing. The design compared three settings: (1) housing projects that had newly established Boys and Girls Clubs with the complete drug abuse prevention program known as SMART Moves; (2) existing clubs in projects that may or may not have had a comprehensive drug abuse program (other than SMART Moves); and (3) housing projects that had no Boys and Girls Clubs. Five new clubs each with the SMART Moves program were assigned two control sites: one public housing site with a Boys and Girls Club without SMART Moves and one public housing site without a Boys and Girls Club. To the extent possible, the control sites were geographically and demographically matched with the experimental sites.

An outside evaluation team composed of evaluators from Columbia University and the American Health Foundation assisted B&GCA with this assessment process. Using a standard interview protocol, members of the evaluation team polled local community leaders, housing authority administrators and residents, and school and police officials to learn the extent of problems and the effects of Boys and Girls Clubs on youth in public housing.

Evaluators examined crime statistics in each site. They also conducted observations, noting the presence of graffiti, garbage, vandalism, drug-related paraphernalia and incidents observed of drug dealing. Through interviews, police officers and community leaders helped to interpret statistics and assisted evaluators in explaining changes that occurred throughout the evaluation.

Specific Outcome Measures:

The project used unobtrusive measures--measures that required little or no participation by the staff and youth, and also used a design characterized by repeated measurements over time. Specific measures were developed for:

- substance use (measured by discarded containers and drug paraphernalia)
- parental involvement with the maintenance of the housing project--for example tenant associations; parent involvement with young people in the youth organizations; and parental involvement in school.
- vandalism and graffiti in unoccupied housing units
- juvenile crimes
- school performance.

Several of these measures were collected with the community rather than the individual participant as the unit of analysis.

Match between Evaluation Design and Underlying Assumptions of the Organization:

Until this evaluation Roxanne Spillett of B&GA had never been fully satisfied with an evaluation design or its results. There were always problems in the implementation of the evaluation, for example, the sample size was too small, program staff lacked the skills to administer surveys appropriately, or participating youth could not be retained over time. Given this history of dissatisfaction, Ms. Spillett laid out the following non-negotiable constraints as she met with potential evaluators:

- Since B&GCA is in the business of serving youth rather than doing research, the evaluation could not disrupt the operations of the Clubs.
- The evaluation had to meet the guidelines established by the funder (OSAP), which included a focus on high-risk youth and a focus on substance abuse prevention.
- The evaluation had to include both process and outcome measures.
- The design had to address the following questions that were of particular importance to B&GCA: What is the impact of a typical Boys and Girls Club newly established in a public housing project? What is the impact of a newly established Club that also provided a comprehensive drug abuse prevention program?
- Finally, the evaluation had to be national in scope and carried out with a limited budget.

Steven Schinke, an evaluator from the Columbia University School of Social Work, was willing to accept the constraints imposed by both B&GCA and OSAP. At the time Dr. Schinke consulted colleague Tom Cook, who recommended that B&GCA not attempt a comprehensive definitive study given the limitations of the setting and budget. So together with B&GCA, Dr. Schinke crafted an innovative evaluation that looked at the impact of the program on the public housing communities as well as on the youth themselves.

Findings:

At the process level B&GCA had wanted to see if the Clubs were actually involving youth in healthy and constructive educational, social and recreational activities. The evaluation demonstrated that housing projects that had Boys and Girls Clubs compared to those that did not had much higher levels of youth program activities.

Furthermore, with respect to planned outcomes:

- For adults and youth alike, there were lower rates of alcohol and drug use, drug trafficking, and other drug related criminal activity in the 10 facilities that had Clubs compared to the five that did not. Housing projects with Clubs were estimated to have 13% fewer juvenile crimes, 22% less drug activity and 25% less crack presence.

- Compared with parents in public housing sites that do not have Club programs and facilities, adult family members in projects with Boys and Girls Clubs were more involved in youth-oriented activities and school programs.

- The presence of crack cocaine and the rates of drug dealing were lowest in sites with Boys and Girls Clubs that had the SMART Moves program. However, differences between the housing projects with Clubs that had SMART Moves and those with Clubs that did not have SMART Moves were not statistically significant.

Fulfillment of Expectations:

Roxanne Spillett indicated that the evaluation very much met the needs of B&GCA. In fact, their funder, OSAP, was so pleased that it presented an exemplary program award to B&GCA. These evaluation results have also:

- led to positive attention from the public, other human service organizations and the media. Business Week magazine published a story on the war on drugs that featured several of the sites from this evaluation.

- positioned the national organization and local affiliates to gain entry into local housing authorities. More than one hundred Clubs have been opened in public housing complexes since the evaluation.

- helped generate funds -- over five million dollars have been raised to support the establishment of Boys and Girls Clubs in local housing authorities.

Lessons Learned:

- There is a need to look beyond the traditional way of approaching evaluation in youth development programs. It is not always appropriate to attempt a comprehensive definitive study, especially when there are programmatic and financial constraints. A better guideline is to implement a well-planned evaluation of a few sites that fits within the constraints of the project.

- In this case it was advantageous to look at the impact of the program not only on the individual participant, but also on the community. The establishment of a Boys and Girls Club empowered parents and citizens, involved the community and had a real impact on many different programs and institutions that affect children's lives.

Highlights from the Girls Incorporated Evaluation of Its Friendly PEERsuasion Program

Girls Incorporated is a building-based youth organization that provides programmatic services to girls and young women. Its mission is to build girls' capacity for confident and responsible adulthood, economic independence and personal fulfillment. There are 200+ affiliates across the country serving more than a quarter of a million young people ages 6-18, most of whom are girls. Each Girls Incorporated facility is professionally staffed and individually run in a way tailored to the needs of the surrounding area. According to Heather Johnston Nicholson, a common organizational saying is, "Once you've seen one Girls Incorporated Center, you've seen one Girls Incorporated Center." The organization defines itself as a provider of informal education rather than recreation. Its core areas of programming include: (1) leadership and community action, (2) health and sexuality, (3) sports and adventure, (4) culture and heritage, (5) self-reliance and life skills and (6) careers and life planning.

Friendly PEERsuasion is a drug abuse prevention program that targets girls from ages 11 to 14. It utilizes a peer education model that recognizes the potentially positive or negative influence of role models on the use of harmful substances by this young population. The program aims to delay the onset of the use of licit substances such as cigarettes, alcohol, and over-the-counter drugs, as well as marijuana and other illicit substances. To accomplish this, the program's fourteen weekly sessions teach participants communication and leadership skills, stress management, coping strategies to deal with peer pressure to use substances, and facts about the harmful effects of using these substances. Following these sessions, girls prepare and conduct similar sessions for younger children from ages six to ten.

Evaluation Design and Outcome Measures:

Girls Incorporated asked Abt Associates, Inc., located in Cambridge, Massachusetts to evaluate Friendly PEERsuasion in four sites around the country. Marcia Chaiken, formerly of Abt Associates and currently with LINC, served as principal investigator. Funding was provided through a grant from the Office for Substance Abuse Prevention (OSAP). The evaluation, utilizing both quantitative and qualitative methods, measured the effectiveness of the program by examining its impact on participants' initiation of substance use, peer associations and coping strategies when confronted with a situation in which harmful substances were present. During the 1988-89 school year, data for the impact study was collected from 354 11- to 14-year-old girls. Due to budget constraints, the process evaluation was conducted in only one of the sites, Birmingham, Alabama, involving 127 girls.

Girls Incorporated wanted to test the generalizability of Friendly PEERsuasion with populations of differing racial, ethnic and socioeconomic backgrounds. Each site had a randomly assigned experimental group that received the program and a control group that was a delayed experimental group, meaning that members of the control groups were able to participate in the program after the evaluation was completed beginning in January 1989. Two of the sites implemented the program with girls in a neighborhood school during the after school hours. The third site offered the program to girls in neighborhood schools during the school day. The fourth site offered the program in four of its centers.

The outcome evaluation was based on self-reports of girls in the experimental and control groups using a repeated measures design. The first questionnaire was administered in September before random assignment occurred and was followed with three self-administered questionnaires in November, January and May. Since the major outcomes of interest were behaviors, the instruments included questions about the girls' behaviors. The evaluation team identified specific questions by networking with other researchers and evaluators such as Gil Botvin and Del Elliott who were generous in sharing their instruments. After the questions were chosen or drafted, evaluators pretested them with 11- to 14-year-old girls (many of whom were slow readers) from another Girls Incorporated center that was not involved in the study.

The process evaluation included inventories which were completed by program participants, parents, and facilitators as well as observations conducted by the researchers. At the end of each session, participants completed a form that asked them if they had participated in specific activities, whether or not they enjoyed the activity and whether or not they learned from them. This gave Girls Incorporated feedback on which activities had been completed and how much the girls enjoyed and learned from them. Facilitators completed forms that indicated who had attended the session and rated each participant's attitude during the session. In addition, Dr. Chaiken made three site visits during which she observed sessions in action and interviewed key people such as program participants, staff from the schools, staff from Girls Incorporated, police and other professionals from the community.

Strengths and Limitations of the Setting:

Girls Incorporated reported many strengths and a few limitations for this type of evaluation.

- Girls Incorporated national and local staff were knowledgeable about and experienced with evaluation. According to Dr. Chaiken, this increased the staff's commitment to the evaluation and enabled staff at both levels to collaborate on the design. It also enabled the local staff to provide good day-to-day oversight of the program delivery and evaluation process within the organization.
- The goals of Friendly PEERsuasion were realistic, well-defined, and capable of being measured.
- The Friendly PEERsuasion curriculum, which was developed and field tested in several centers in Texas, is well-designed and thorough. Girls Incorporated knew what was supposed to happen and could, therefore, monitor program delivery during the project study. Evaluators knew with confidence that the trained peer leaders were going through the same 14 weeks of activities in all of the sites. However, it was less clear what the peer educators would actually do after their training.
- The program was very appealing to participants. This provided evaluators with a large enough group of participants to randomly assign them to an experimental or control group. As mentioned earlier, the control group was actually a delayed participation group. Girls Incorporated, like most youth development organizations, was not willing to withhold an important program from any girl in the community who wanted it.
- The primary study site, Girls Incorporated of Central Alabama (Birmingham), had a 50-year history of developing programs and delivering them to girls in that community. So evaluators felt confident that they were assessing the program itself rather than strengths or weaknesses in the organization.
- In Birmingham, Friendly PEERsuasion was delivered in the schools. Often problems arise when one organization delivers a program on another organization's turf. The fact that Girls Incorporated of Central Alabama had the support of the superintendent and the board of education was a definite strength.
- The primary limitation was a lack of funding. The program was being implemented in four sites, but Girls Incorporated

had funding for evaluation in only one site. This limited the process evaluation to only one of the sites. In spite of insufficient funds, Girls Incorporated collected data for the outcome evaluation in all four sites and eventually got an additional private grant to analyze all of the data.

- It was challenging for Girls Incorporated to obtain access to large enough numbers of program participants to carry out a random assignment design. In order to get at least 100 participants, three of the sites found it necessary to take the program into the schools. As stated earlier, the Birmingham site, having a strong relationship with the schools, was able to offer the program during the school day. The other two sites, however, had only limited relationships that they were trying to strengthen through this project. They offered Friendly PEERsuasion as an after-school program. Because the relationships were new, and because the local affiliates didn't have enough leverage in that context, they encountered significant problems with participant attrition.

Match between the Evaluation and the Underlying Assumptions of the Organization:

According to both Dr. Chaiken and Dr. Nicholson, the match was very appropriate. Abt Associates actively involved Girls Incorporated staff in the design of the overall evaluation and questionnaires. The evaluation also attempted to incorporate some of the broader philosophies of the organization. For example, Girls Incorporated believes strongly that girls are decision-makers and leaders and that girls should be actively involved in their learning. In keeping with this ideal, evaluators treated the girls with respect, educated them about the evaluation process and involved them in the actual implementation of the evaluation. The girls learned about random assignment and pretesting. In the process evaluation one girl in each program section was assigned the role of research assistant. These girls wore badges labeled PEERsuader Research Assistant, passed out the questionnaires, made sure that they were complete, collected the questionnaires, put them in an envelope, sealed it, and sent the questionnaires to the evaluators for analysis.

There were still problems, however, with using a quasi-experimental design in a youth organization. The three primary challenges focused on: (1) ethical concerns about random assignment, (2) unintentional errors made by program staff and (3) the volume of paperwork associated with data collection.

All staff -- national and local -- within Girls Incorporated were concerned about the ethics of random assignment. In fact the organizational guidelines state that evaluations should avoid random assignment. However, the internal review board for this study felt reasonably comfortable with this particular design, a delayed entry control group, that involved withholding the program for only one academic semester.

Although the local staff were knowledgeable about evaluation and enthusiastic about the project, some of them took actions at the program level that negatively impacted the evaluation. For example, some local staff changed the program as they delivered it, helped girls complete questionnaires, asked intrusive questions, or offered another drug prevention program in addition to Friendly PEERsuasion in their site during the time of the study.

Finally, even though many steps were taken to reduce the burden of data collection, the standard operating procedures of the Girls Incorporated centers did not lend themselves easily to the vast amounts of required paperwork. Both staff and girls were frustrated by the forms they had to complete over and over.

Findings:

- The girls' background characteristics predisposed them to risk of substance abuse. For example, 89% qualified for free lunch, 25% aged 12 or younger were unsupervised at home after school and 44% had mothers who smoked cigarettes.

- Friendly PEERsuasion was popular with the girls. Attendance was high and girls reported liking over 93% of the activities.

- Early adolescence is a critical period in the transition from nonuse to use of harmful substances. A majority of the girls ages 13 and 14 had already tried smoking, drinking or using other drugs at the beginning of the study. For younger girls (ages 11-12), this was a critical period in the transitional process from non-use to use of harmful substances.

- The program had a discernible but not dramatic effect in delaying the onset of substance use. At the end of the program, girls who had participated were less likely to have initiated use of harmful substances than those in the control group. Even within a few weeks after the girls finished the program, the effects started to wane. Thus, short term programs are likely to have short term effects.

Given these findings, Girls Incorporated offers the following recommendations: (1) focus prevention on preadolescents or the youngest adolescents, and (2) aim for continuity in programs throughout the teen years.

Lessons Learned:

- In addition to more traditional methods such as questionnaires, one of the best ways to get data for a process evaluation is to talk to participating youth themselves.

- Much important data was gleaned from the process evaluation that was conducted in only one of the sites. It would have been especially helpful to have that kind of observation in all four sites. The process evaluations revealed, for example, that PEERsuaders were patterning their language, their gestures, and their whole concept of teaching specifically on the adult who was leading them. This finding gives credence to the belief that peer educators can learn behaviors that will enable them to be positive role models for other girls. This type of data is difficult to retrieve from checklists and questionnaires. It is critical that evaluations of youth development programs go beyond the effect/no-effect approach in trying to figure out what is working for whom and why.

- The follow-up analysis of all four sites assessed the impact of the program across all four sites. This analysis would have yielded more if the study had been partitioned site by site but the size of the data set constrained the extent of site by site analysis.

- Evaluators may have underestimated the level of information that Girls Incorporated staff needed to effectively support the evaluation at the local level.

- Having a delayed-participation group within the same site as the experimental group did lead to some contamination. Evaluators attempted to control for this by asking the delayed group if they knew anybody who had participated before. This process showed that girls who had delayed participation were in fact getting some of the same positive attention that went along with being a part of a national study as members of the experimental program. They knew they were going to eventually be PEERsuaders. Each week girls in the delayed group approached the staff to find out what was going to happen.

- Participating in this type of serious evaluation appeared to increase the commitment of the local staff, who were already taking their work with youth very seriously.

- In an attempt to produce meaningful results, evaluators are often too rigorous in the questions they ask in a quantitative evaluation study. They must be careful to make allowances for constraints within youth development organizations that try to do this kind of systematic assessment. Evaluators also have to make allowances in interpretation so that, for example, a 10% confidence level may be more appropriate in testing the impact of a program for youth than it would be in testing a new cancer drug. There is also a need to reinterpret what is meant by terms such as rigorous and measurable effects.

SEGMENT THREE: PARTICIPANT DISCUSSION

The afternoon session was devoted to open discussion in particular about evaluation of youth development programs at the local level. The following, sometimes divergent, viewpoints were expressed about the need for and ability of local organizations to conduct meaningful evaluations:

- Local organizations need assistance in order to conduct their own evaluations. Participants suggested providing training -- workshops, summer institutes, and so on, compiling handbooks that include measurement instruments that people can use, and offering technical assistance.

- There is a need to address the cultural sensitivity of evaluation approaches and measures at the local level. We must adapt some instruments to make them relevant to various racial and ethnic groups as well as identify, recruit and train students of color who might specialize in evaluation of youth development programs in the future. Local organizations should also attempt to identify evaluators who match the demographics of their specific communities.

- Several participants voiced reservations about pushing local organizations to engage in rigorous outcome evaluation. Many local organizations do not have the capacity for such evaluation efforts and would be diverting energy and resources that would be better used for providing services to children directly. Ideally, the role of the national organizations is to conduct a rigorous outcome evaluation that can then be used by their local affiliates, assuming the local agencies apply the essential principles that were the basis for the evaluation. National organizations are more likely to have the resources and access to multiple sites necessary to carry out expensive impact evaluation studies.

- United Way is often criticized for imposing evaluation criteria on local organizations. Russy Summariwalla and Martha Taylor of United Way of America discussed a new initiative being undertaken that aims to build the capacity of indigenous organizations to conduct their own local evaluations. This initiative will provide local organizations with the tools, expertise and technical assistance needed to do program evaluation.

- Evaluation is essential at two different levels. One level is the kind of rigorous evaluation that is needed to document whether a certain type of intervention has a certain type of effect. Typically this type of evaluation is done at the national level. Then local organizations can pick up those tested interventions, making the assumption that they will work when implemented at the local level. There is also a need, however, for evaluation at the second level. Local organizations also need to document the immediate results of their interventions with individuals on a day-to-day basis. For example, what is the result of a one year experience in Boy Scouts? The United Way would like to facilitate the capacity of local organizations to conduct this type of evaluation.

- Funders often ask unrealistic questions of the organizations they are funding and expect unrealistic results. Agencies must educate funders and decision-makers about the amount and type of documentation they request from service providers.

- P/PV has found it helpful to do a uniform management information system (MIS) across the various youth development projects they have evaluated. At this point P/PV has compiled a small set of key variables that both identify the population and describe service delivery. Whenever it enters into a new relationship with an organization, P/PV asks the staff to customize the measures for their interest. Once P/PV discontinues its work with that organization, the local staff are left with increased capability to handle evaluations. Though no local organization that has used this process necessarily likes it, by the end of the process they have a better appreciation of what the data can do for them as well as P/PV.

- Local evaluation can answer some questions very well and other questions very poorly. Local evaluations should be rigorous enough to answer six questions: (1) Does this study let me see how the program is really implemented?; (2) Does this study clearly reveal the outcomes of the intervention?; (3) Will I be able to say that the program caused or led to the outcomes I've found?; (4) Does this design let me see reasons for varying amounts of success?; (5) Does this study give me information for extending my program?; and (6) Are the operations of the study generally replicable?

- Staff in local organizations need training that will help them develop the kinds of thinking, documenting, and reporting that will allow them to tell whether they are accomplishing their goals even if they cannot rule out all other factors as possible causes of positive results.

- Local evaluation needs to be much more program developer centered and indigenous to the program, and much less dependent on an outsider coming in with predetermined standards.

- What is the distinction between local evaluation and national evaluation? What is the distinction between basic research and evaluation research? What is the distinction between process evaluation and outcome evaluation? Are we doing the field a disservice by making distinctions that do not get at the heart of what is happening with youth?

- One of the difficulties with program evaluation in this particular field is that it is often simultaneously trying to do basic research and program evaluation. The field of youth employment, for example, does not have the burden of having to demonstrate that having a GED or a high school diploma makes a person more employable. But there is no basic research that says participation in a youth program leads to any specific outcomes. So local organizations are left to try to document that a particular experience they provide is associated with a particular outcome and that the organization did a good job of providing that experience.

- There are hundreds of free-standing youth organizations that are not connected to one another. Perhaps there should be a long-range strategy to connect them to each other, to third parties like P/PV and to evaluators in local universities. In the meantime there are some national delivery systems that offer many advantages to program evaluation efforts. National evaluations provide (1) an opportunity to test a program in diverse sites that may be representative of the nation and (2) a mechanism for feeding the effective program out to affiliates for replication. The best national evaluations are partnerships between national and local organizations.

- Replicability in diverse settings is not always a necessity for a program to be judged useful.

Near the end of this segment participants briefly addressed the question: How much weight should be placed on evaluating the impact of an organization's core programming vs. categorical programming? Participants who responded stated:

- It is more important to evaluate the core program, the very essence of the organization and the difference it makes for the lives of young people. Evaluations of categorical programs are also important but present a lower priority to many organizations. Unfortunately, there is usually more funding available to evaluate special projects and categorical programs.

SEGMENT FOUR: FUNDERS PANEL

In the fourth segment three panelists (listed below) presented important views on the roles and perspectives of funders with regard to program evaluation within youth development organizations.

Donna Dunlop, Program Director, DeWitt Wallace-Reader's Digest Fund

Gloria Primm Brown, Program Officer, Carnegie Corporation of New York

Hector Sanchez, Public Health Advisor, Office for Substance Abuse Prevention, Department of Health and Human Services

The panelists were asked to address the following questions in a 15-minute presentation:

a. How important is an evaluation component in determining whether a new program is funded? What kinds of evaluation designs are most impressive to you? How important is cost as a factor in assessing a particular evaluation design?

b. What strategies does your foundation or agency use to follow up on a grantee's evaluation process?

c. How do you use program evaluation results? How do evaluation results impact future funding of an existing program? new programs being proposed? What is the role of the funder in disseminating evaluation results?

d. How strong must evaluation results be for you to consider a program effective?

Each of the three panelists offered a unique perspective in responding to these questions. Their responses are summarized below:

Importance of Evaluation

- Program evaluation has become increasingly important to the DeWitt Wallace-Reader's Digest Fund, which is a fairly new funder of youth development agencies. Four years ago the board was interested in funding programs that "have an impact on kids," without recognizing the need to make an investment in the documentation and evaluation of those programs. Now the Fund is financing evaluation for two primary reasons. First, it helps the board document whether the Fund's investments in programs are making a difference for youth. Second, it helps grantees to be more clear in identifying their goals and marking their progress toward reaching those goals.

- The DeWitt Wallace-Reader's Digest Fund considers the following criteria when reviewing a proposal: What is the program? What needs will it meet? What staff will deliver it? Is it local or national? How will the program document that it is achieving its stated goals? More often than not when prospective grantees approach the Fund, they do not have a plan for program evaluation. If the Fund is interested in a new program, the program officer will begin a conversation or a negotiation about what is needed. The Fund expects that an existing program will have an evaluation component, but will not necessarily withhold funding for an untested program.

- The Carnegie Corporation tends to support new projects and prefers to support national organizations. Carnegie has found that many youth-serving organizations paid little attention to evaluation until their funders began raising questions about outcomes. The Corporation reviews hundreds of proposals over the course of a year, and sometimes those proposals are quite similar. Therefore, staff must gather information to best determine which programs should be funded.

- For its High-Risk Youth Demonstration Project, OSAP has specific requirements for program evaluation efforts. All prospective grantees are expected to include an evaluation component in their proposals that addresses the following four areas: (1) review of the overall picture -- who, what, why and how; (2) description and evaluation of program activities; (3) evaluation of administration and staffing; and (4) outcome evaluation. Each project is expected to engage in both process and outcome evaluation activities.

- There are no hard and fast rules about the funding of program evaluations. Some funders require them, others do not. Some foundations can only support the service delivery component of a project. There are a few foundations that can only fund evaluations. And then there are some that can support both. All foundations have constraints that are dictated by their charters and by their boards.

- Foundations differ from other kinds of funding agencies and from one another. One distinction is whether a foundation is local or national. Large national foundations such as Carnegie, Ford, and Rockefeller seek to build new knowledge and, therefore, tend to be oriented to research and evaluation. These national foundations expect prospective grantees to have given thought to how they will assess what they are proposing to do.

- Many grantees seem to feel threatened by a funder asking for documentation of a program's effectiveness. Agencies must begin to realize that it is important to gather this information for their own sake. Practitioners need to evaluate in order to make appropriate changes to improve the program, to learn about what works and why and to make a case for raising money from other funders. Funders appreciate grantees who recognize that they need to be undertaking evaluation for themselves.

- In some foundations there is tension between the board and the staff about program evaluation. Staff tend to be more familiar with the dynamics of programming and are more likely to promote evaluation that brings in outside evaluators who will work collaboratively with the grantee. Board members are sometimes more inclined to believe that evaluation should be planned and implemented by objective outsiders with little or no input from the grantee.

Most Effective Evaluations

- Carnegie likes to see attention to evaluation in every project that it funds, but the foundation's expectations vary depending on the proposed project. A large demonstration project that provides services would not be funded without an evaluation component. Carnegie often sends proposals to outside experts to get their feedback not only on the quality of the proposal but also on the evaluation design. Feedback from the experts is then passed on in an anonymous fashion to the prospective grantee. If the criticism is constructive and valid, Carnegie expects the prospective grantee to make changes to strengthen the project and the evaluation. On rare occasions, Carnegie has even gone so far as to set up an evaluation committee for the grantee and to work with them to improve and strengthen their program. More often, the foundation helps prospective grantees identify consultants or experts who can help them clarify their proposed plan.

- The DeWitt Wallace-Reader's Digest Fund likes evaluations that are realistic. Too many evaluations are overly-ambitious from the start. Staff-driven, staff-involved evaluations are very important to the Fund. They want the simple questions answered first before attacking the more complex questions. If a project has multiple sites, the Fund is interested in whether there are mechanisms in place to gather information across sites. The Fund also looks for a balance between qualitative and quantitative data, having been persuaded by both types.

- Whether or not an evaluation may be required, or what kind of evaluation, how rigorous it is, the types of measures used, will depend on the project's goals, the population to be served, the cost, and why the funding agency decides to fund it. To illustrate the last point, a corporation with offices in a certain locale may be motivated to fund an interesting project because of the good public relations that will accrue to the corporation. And that corporation may be content with anecdotal information about the quality of the grantee's services, and never require any kind of outcome evaluation. Another foundation or corporation may be interested in fostering leadership development among minority organizations, and therefore may place more emphasis on service delivery than on the evaluation component.

Follow-Up Strategies/Uses for Evaluation Results/Role of Funders in Disseminating Findings

- Most funders request an annual report that serves as a vehicle for keeping the funder abreast of the grantee's ability to carry out the evaluation as planned. Foundation staff also sit on advisory committees and sometimes hire outside consultants to review reports and to complete summaries.

- Most funders use evaluation results to help determine whether to fund other similar projects.

- The DeWitt Wallace-Reader's Digest Fund will help to disseminate evaluation reports if they will serve the needs of the grantees. Some evaluations that were not well thought out or implemented have been released to the public and, quite naturally, have resulted in damage to the programs. So funders need to be responsible in deciding how an evaluation should be interpreted and disseminated.

- All projects that receive grants from OSAP have to participate in a national evaluation plan. OSAP also requires quarterly progress reports for the first year. OSAP publicizes impressive evaluation results in a bimonthly information service, the Prevention Pipeline. This publication also contains updates on OSAP's program activities, news about prevention efforts at the federal, state and local levels, tips for getting prevention stories in the news, summaries of research findings that have immediate program application, and abstracts of key research findings.

- Sometimes OSAP can obtain funds from the Government Printing Office to print and distribute evaluation reports through the National Clearinghouse for Alcohol and Drug Information. OSAP also shares the final reports from all of its grantees with ERIC and with Project Share.
- The Carnegie Corporation encourages grantees to disseminate findings through books, monographs, newsletters, professional journals, other forms of media as appropriate such as videotapes, and by appearing on panels at professional meetings or at meetings of advocacy organizations. The Corporation also encourages and supports multi-disciplinary meetings between researchers and practitioners and briefings with legislators and policymakers to report important program findings. A few programs have their findings disseminated through the Carnegie Quarterly.
- On occasion, Carnegie will provide funding to disseminate an evaluation report that it did not originally fund if the results are powerful.

Suggested Resources

- A publication from the National Research Council, Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing, which was published in 1987, offers helpful information about elements of successful programs and program evaluation. The panel on Adolescent Pregnancy and Childbearing, which contributed recommendations for Risking the Future, concluded that every program may NOT be worthy of a formal evaluation (since the evaluation can sometimes cost as much as the program itself). The panel also recommended that federal and state funding agencies set aside support for evaluation research, and that the research community take an active role in designing and helping programs to design and implement these studies.
- The Handbook for Evaluating Drug and Alcohol Prevention Programs, which is available from the National Clearinghouse on Alcohol and Drug Information, also contains helpful ideas for youth development agencies.

SEGMENT FIVE: SMALL GROUP DISCUSSION OF RECOMMENDATIONS

In the final segment participants divided into small groups to identify their three most important recommendations for strengthening program evaluation efforts within youth organizations. The range of participants from youth development specialists to evaluation experts was represented in each group. The recommendations from each of the three small groups follow.

Recommendations from Group One:

1. Make a better case for evaluation at the local level, articulating the salient questions that are reasonable to expect local evaluations to answer well. See evaluations conducted by hands-on youthworkers as a way to ease burnout.

2. Increase the focus on outcome evaluation within the youth development field at both the national and local levels. Local agencies will need to look outside their own organizations, perhaps to local universities, for the technical help they need. Ideally, university evaluators could provide local agencies with instruments that have been used successfully in other studies. Consider locating individuals trained in ethnographic or qualitative evaluation in community organizations for three weeks or more of on-site observation. This recommendation requires funding to compensate local universities for their technical help.

3. Place greater emphasis in national evaluations on the assessment of:
   - programmatic impact -- especially functional long term outcomes.
   - the quality of program implementation.
   - processes surrounding the operation of programs -- both theoretical and practical.
   - the historical, political, and human context in which programs are embedded.

4. Utilize a rigorous analysis to identify the outcome measures for youth development studies. There is a need to look behind the slogans of the day as in the example of self-esteem, which neither in substantive theory in psychology nor in terms of direct policy outcomes can be justified as a major outcome measure. Whenever possible, choose functional outcome indicators such as those utilized in the evaluations of the Head Start program. Head Start was a political success because it demonstrated that over the long term participants were performing better in the labor force and had more intact families, both of which are part of the indicators of successful entry into the adult world.

Consider studies of the alumni or past participants in programs.

Recommendations from Group Two:

1. Set up a national resource center/network, based on the following criteria:
   - it must be permanent;
   - it should be free-standing--not connected to any single funding institution or organization;
   - it should be action-oriented rather than merely a repository of information; and
   - it must provide services to both national organizations and local organizations.

   The mandate of this center will need to be further refined but it must play multiple roles that include being a repository for both primary research and program evaluations, synthesizing evaluation data, offering technical assistance, and perhaps, linking local organizations with experienced evaluators.

   It would be advantageous to look to the resources in the Youth Development Information Center, currently located in the National Agricultural Library, as a beginning point for this project.

2. Create a pool of resources, both monetary and technical, to encourage a stronger focus on evaluation, especially at the local level. Bring together people who have some expertise in evaluation, whether they're national organizations or independent researchers, and provide them with an incentive to share that information. Funding might come from an interdisciplinary set of sources such as national foundations, community foundations, as well as the federal and state governments.

3. Convene an interdisciplinary team to discuss youth development. Ideally, this group would move forward into evaluating efforts to promote positive youth development rather than evaluating success at damage control. Once this team has reached some conclusions, recommendations could be fed back into the national center, to funders and to the academic and research communities.

Recommendations from Group Three:

1. Establish a technical assistance/national resource center that would:

   - synthesize existing research.
   - publish a newsletter.

- Create a nationally accepted Management Information System for the field of youth development.
- Provide training and staff development.
- Provide technical assistance.
- Refer agencies to independent evaluators and to personnel at local universities.

This center could be created as a result of a partnership between researchers, youth organizations and funders. Perhaps the major national organizations could each contribute a certain amount of money annually to help fund this center. Shared funding would increase the commitment of all the stakeholders in evaluation efforts.

2. Identify a credible range of youth development outcomes and indicators for those outcomes. One strategy for identifying outcomes is to conduct an intensive case study with youth to find out from their perspective what actually goes into a youth development program, and what the youth feel they are getting out of it.

The chair, Dr. Torney-Purta, noted the similarities in the three sets of recommendations. She commented on the way in which all groups focused on creating both a vision and the will to bring an infrastructure into being to work across organizations and across program experts and evaluators to improve evaluations in the youth development field. Jane Quinn told participants that their ideas would be integrated in to the final report on the Task Force on Youth Development and Community Programs, and that they would receive a written report on evaluation of youth development programs, including a summary of today's proceedings. In closing, Dr. Torney-Purta invited participants to send her any additional thoughts about the meeting and thanked the group for their participation.

Postscript: Following the meeting, Marcia Chaiken of LINC sent in writing the following reflections about recommendations made during the Consultation:

- If evaluations focus exclusively on "functional" behaviors, we will ignore the risk-taking aspects of adolescent development and will not learn how youth programs can effectively channel this developmental process into productive avenues. If we focus exclusively on long-term outcomes, we will be hard pressed to explain the outcomes, for example, to understand why some participants eventually become gainfully employed adults in stable marriages who contribute to their communities, while others manifest socially undesirable behaviors. It will also be difficult to demonstrate that the program actually had something to do with these outcomes.

- Focusing on indirect effects on the community rather than direct effects on participants' immediate behavior will lead to questions of whether the outcomes were actually caused by the program. The history of evaluation research suggests it will be best to evaluate specific direct effects youth programs are designed to achieve, and use outcome measures documented as part of basic research about adolescent development.

- We must think carefully about promoting a clearinghouse for information about youth programs and evaluation findings. In this modern information world, clearinghouses can be just one more intermediate step in locating needed material, and they can even become a bottleneck to research and program development if understaffed. A better strategy might be to adapt a model -- a publication for District Attorneys and their staffs -- developed by Steve Goldsmith, mayor of Indianapolis. This model publishes regular digests of studies and evaluations of DA's practices together with practical critiques written by respected and influential District Attorneys.

Attachments:

Appendix A: Roster of Consultation Participants
Appendix B: Summary of Participants' Written Comments
Appendix C: Summary of the State of Program Evaluation Within Selected National Youth Organizations
Appendix D: Summaries of Selected Articles
Appendix E: Bibliography of Evaluation of Youth Development Programs

APPENDIX A

CARNEGIE COUNCIL ON ADOLESCENT DEVELOPMENT
CARNEGIE CORPORATION OF NEW YORK

CONSULTATION ON EVALUATION OF YOUTH DEVELOPMENT PROGRAMS
FINAL PARTICIPANT LIST
January 15, 1992

Meeting Location: Staff and Board Room
26th Floor
Carnegie Corporation of New York
437 Madison Avenue (at 50th Street)
New York, New York

CHAIR

Judith Torney-Purta
Professor, Department of Human Development
Benjamin Building, Room 3304S
University of Maryland
College Park, MD 20742-1131
(301) 405-2806

PARTICIPANTS

Alvia Branch
Vice President and Director of Qualitative Research
Public/Private Ventures
399 Market Street
Philadelphia, PA 19106-2178
(215) 592-9099

Sylvia Brown
Director of Research
Girl Scouts of the U.S.A.
830 Third Avenue
New York, NY 10022
(212) 940-7500

Marcia Chaiken
Director of Research
LINC
Post Office Box 406
Lincoln, MA 01773
(617) 259-8222

Thomas D. Cook
Professor
Center for Urban Affairs
Northwestern University
2040 Sheridan Road
Evanston, Illinois 60208
(708) 491-4900

Donna V. Dunlop
Program Director
DeWitt Wallace-Reader's Digest Fund
261 Madison Avenue, 24th Floor
New York, NY 10016
(212) 953-1201

Larry Eisenberg
Director of Research and
Strategic Planning
Boy Scouts of America
1325 West Walnut Hill Lane
Post Office Box 152079
Irving, TX 75015-2079
(214) 580-2000

S. Shirley Feldman
Professor
Stanford Center for the Study of Families, Children,
and Youth
Stanford University
Building 460, Room 150
Stanford, CA 94305-2135
(415) 725-2495

Karen Fulbright
Program Officer
Ford Foundation
320 East 43rd Street
New York, NY 10017
(212) 573-5000

Dagmar McGill
Associate Executive Director
Big Brothers/Big Sisters of America
230 North 13th Street
Philadelphia, PA 19107
(215) 567-7000

Matthew Miles
Senior Research Associate
Center for Policy Research
475 Riverside Drive
Suite 7228
New York, New York 10015
(212) 870-3186

Mirka Negroni
National Director for Youth Services
The ASPIRA Association
1112 16th Street, N.W.
Suite 340
Washington, D.C. 20036
(202) 835-3600

Heather Johnston Nicholson
Director
National Resource Center
Girls Incorporated
441 West Michigan Street
Indianapolis, IN 46202
(317)634-7546

Karen Pittman
Director
Center for Youth Development and Policy Research
Academy for Educational Development
1255 23rd Street, N.W.
Suite 400
Washington, D.C. 20037
(202) 862-1842

Hector Sanchez
Public Health Advisor
Office for Substance Abuse Prevention
5515 Security Lane, 9th Floor
Rockville, MD 20852-5003
(301) 443-0353

Steven Schinke
Professor
Columbia University School of Social Work
622 West 113 Street
New York, NY 10025
(212) 854-8506

Roxanne Spillett
Assistant National Director of Program Services
Boys and Girls Clubs of America
771 First Avenue
New York, NY 10017
(212) 351-5900

Russy Sumariwalla
Senior Consultant
United Way Strategic Institute
United Way of America
701 North Fairfax Street
Alexandria, VA 22314-2045
(703) 836-7100

Billie Tisch
980 Fifth Avenue
Apt. 25B
New York, NY 10021
(212) 249-6441

Renee Wilson-Brewer
Associate Director
Center for Health Promotion and Education
Education Development Center, Inc.
55 Chapel Street
Newton, MA 02160
(617) 969-7100

Joan R. Wynn
Research Fellow
Chapin Hall Center for Children
University of Chicago
1155 East 60th Street
Chicago, IL 60637
(312) 753-5900

OBSERVERS

Linetta Gilbert
Program Officer
Greater New Orleans Foundation
2515 Canal Street, Suite 401
New Orleans, LA 70119
(504) 822-4906

Imre Kohn
Director of Evaluation
Milton S. Eisenhower Foundation
48 Charles Street
New York, NY 10014
(212) 924-1435

Martha Taylor
Director, Market Research Products and Services
United Way of America
701 North Fairfax Street
Alexandria, VA 22314
703-836-7100

CARNEGIE CORPORATION OF NEW YORK

Gloria Primm Brown
Program Officer
Carnegie Corporation of New York

Jane Quinn
Project Director
Carnegie Council on Adolescent Development

Vivien Stewart
Program Chair
Carnegie Corporation of New York

Ruby Takanishi
Executive Director
Carnegie Council on Adolescent Development

Pamela Wilson
Consultant
Carnegie Council on Adolescent Development

SUMMARY OF RESPONSES TO QUESTIONS

Participants in the Consultation (for roster of participants, see Appendix A) responded in writing, in advance of the meeting, to three questions: 1) What progress has the field of youth development made in program evaluation over the last 10 years? 2) What are the major barriers to progress in program evaluation within the youth development field? 3) How can we build on and strengthen program evaluation efforts in the next decade?

There were many similarities but also some unique perspectives in participants' responses to these questions. This section will summarize the key issues and recommendations that surfaced.

1. What progress has the field of youth development made in program evaluation over the last 10 years?

Several participants noted that some progress had been made:

a. Prevention programs (adolescent pregnancy, school dropout, and substance abuse) that are an integral component of many youth development efforts have been rigorously evaluated within youth development agencies and have demonstrated results at varying levels. In the specific discipline of health behavior prevention, great progress has been made in program evaluation of cancer, drug abuse and cardiovascular disease prevention. New methodologies have emerged from these studies--sampling, research design and analysis. The science of prevention has emerged and been legitimated.

b. Although little attention has been directed toward systematically and quantitatively assessing the effectiveness, self assessment has continuously taken place. Youth organizations have proven their effectiveness to the youth with whom they work as evidenced by their continued enthusiasm to join, verbal and written statements of satisfaction with the services being provided and through their continued connection with the organization as staff, volunteer, board member, donor, and so on.

c. Because youth development professionals do not typically have the expertise or the funds to systematically measure changes in positive outcomes (such as increased self worth, social acceptance, ability to trust, leadership skills, decision-making skills), they have preferred to spend the time working with kids rather than gathering statistical data to report on their progress. Youth organizations assess progress more by monitoring a young person's achievement in school, enthusiasm and involvement, by knowing the kids and befriending them, by learning what makes them tick from the inside, and less by measuring them "objectively" from the outside.

d. The emerging field of youth development is increasingly aware of the need for and importance of program evaluation. Intermediary organizations such as the Manpower Demonstration Research Corporation (MDRC) and Public/Private Ventures (P/PV) have played an important role in this area.

e. Program administrators at both the national and local levels are cooperating fully in evaluation partially because of requirements imposed by funders and partially because of a sincere interest to learn more about the effects of program activities on participating youth.

f. Youth development organizations with an internal research capacity seem more likely to undertake sophisticated evaluations, even if much of the work is done by an outside research team.

g. There is a salutary trend toward process and outcome evaluation that flows naturally from programs with carefully conceived goals and objectives. Such evaluation is based on good record-keeping and astute observation by the program implementers and participants, sometimes assisted by an independent observer. A few key youth development organizations are sufficiently far along in their capacity to conduct this type of evaluation with support and coordination from a few key funders.

h. In certain areas, evaluation methodology used in program evaluation has been strengthened. The application of experimental and quasi-experimental research methodology to evaluate youth programs, particularly in the area of youth employment, has increased the demand for and acceptability of research findings.

i. Although there is tremendous variation among evaluation consultants working in the youth development field, many evaluators have become more responsive to the concerns of program administrators and implementers. Rather than seeing themselves strictly as disinterested outsiders, evaluators are increasingly serving as members of a project team, communicating more openly with other project team members and making recommendations for improving projected outcomes.

j. Evaluators have had to become more flexible in their evaluation designs and questions, for example by creating designs that anticipate youth attrition in reaction to the reality that even the most appealing youth programs lose participants. Evaluations are more likely now to explore why some youth continue to participate while others drop out. Rather than simply finding out whether a particular program outcome is statistically significant, evaluators are attempting to determine the types of youth who are positively affected by the program and why the program was less effective with other youth.

k. Methodological progress has been made through the use of a wider set of research tools. Evaluators have begun to develop innovative statistical or qualitative techniques and to reach out to colleagues in other disciplines and borrow some of their analytical tools. While some evaluation experts and funders insist on quantifiable results, the importance of context to the success of a program suggests the inclusion of qualitative approaches as well.

l. Theoretical underpinnings of program evaluation and interpretation of findings have been fertilized by cross-disciplinary interactions. For example, in working with "high risk youth," we are more aware that factors that place a young person at higher risk for adolescent pregnancy also place him or her at higher risk of dropping out of school. In addition, with increased knowledge of adolescent development, we are learning that even within a specific program model, certain types of approaches may be highly effective for preteens and not at all effective with older adolescents.

m. There is increasing awareness of how evaluation and research differ from and yet complement one another. Evaluation is a relatively short-term inquiry about a particular program or project designed to inform decision-making, often about the direction of a program or its funding. Research is usually a long-term inquiry related to a theoretical or conceptual framework that attempts to produce credible generalizable knowledge about psychological, educational, or social processes. While our work at the consultation will be informed by knowledge of research on adolescence, the primary focus of the meeting will be on evaluation. We want to know how to make evaluations of youth development programs better serve the needs of the stakeholders in these programs (ranging from the organizations that provide funding to the youth who participate).

n. Lessons from other fields are available to apply to evaluation of youth programs. For example, in the field of global education, a series of activities initiated by the American Forum for Global Education (the development of a compendium of instruments for process and outcome evaluations, a conference on evaluation, technical assistance to aid projects in carrying out their evaluation plans and a set of workshops to train program practitioners in conceptualizing and conducting evaluations of education about international development) led to much more positive attitudes toward evaluation among the program staff.

Other participants were less optimistic:

o. Because there is no codified youth development field, it is difficult to document specific progress. Most importantly, the field has not been uniformly defined and positive outcomes that might be evaluated are only now being identified.

p. Overall, significant gains have not been made in program evaluation, and in fact, youth organizations are increasingly aware of problems and concerns associated with evaluation. However, youth organizations are more aware of:
  - evaluation techniques other than the use of pre/post questionnaires,
  - the use of standards of measurement (e.g., quality program indicators) to establish quality improvement objectives and to measure results, and
  - the use of computer technology to aid in tracking participation of members.

q. The main progress the "field" has made in the past ten years is that it now takes evaluation, particularly outcome evaluation, seriously.

2. What are the major barriers to progress in program evaluation within the youth development field?

a. We don't have a full understanding of the roles youth development programs can play in the lives of young people. Much of our current thinking about program effects is based on common-sense theorizing and generalizations made from poorly-designed evaluations.

b. The criteria for a program's success are elusive, and are not always well specified from the inception of the program. Sometimes effects of a program may be found in unanticipated domains.

c. The global theorizing and findings drawn from existing research in the field may lead to over-promising, inflated claims about what youth development programs can do for participants. If programs are oversold, the results of rigorous evaluations are likely to lead funders and policy-makers to look for a new set of interventions that promise major effects but whose promises have yet to be tested.

d. The cost of evaluations is highly prohibitive in that the types of evaluations that are likely to yield the most reliable information are the most expensive to conduct. Youth organizations lack the financial resources to carry out scientifically rigorous evaluation. Funding is insufficient at all government and private sector levels.

e. There is a general lack of knowledge about evaluation among youth development professionals. Both administrators and practitioners in local organizations: 1) are confused about the different levels and types of evaluation (process vs. outcome), 2) are intimidated by the terms, 3) don't know how to plan or budget for evaluation, 4) perceive evaluation as a burden, a threat or something unnecessary rather than an asset, and/or 5) don't believe that the value of youth development organizations can actually be documented. Training in program evaluation that parallels the organization's mission and goals is rare.

f. Program staff are usually paid only for the time they are with youth, making it difficult for them to plan or evaluate.

g. There is no consistent mechanism for sharing ideas and/or findings and discussing issues related to evaluation in the youth development field.

h. Rigorous evaluations utilizing an experimental or quasi-experimental design can be problematic in youth organizations. Such evaluations impose a heavy burden on the staff and participants of a program and therefore require careful planning to reduce the burden, provide incentives to compensate for the burden, and sustain enthusiasm of all concerned. Examples of problems that commonly occur include:
    - ethical dilemmas related to random assignment of participants to treatment and control groups;
    - special challenges for recruitment and retention of participants given the voluntary nature of activities;
    - small sample sizes;
    - difficulties matching control/comparison and experimental groups.

i. Interventions are rarely simple and often are of the "prevention" variety so that outcomes (pregnancy, drug use, school leaving) are relatively rare events, requiring subject pools that range somewhere between large and enormous to get statistically significant results.

j. Evaluators are often called in too late after program planning is completed. They need to be a part of the planning process.

k. Evaluation methods and results are most likely to be presented to and discussed among researchers rather than among staff developing or implementing youth development programs. An enhanced contribution could be made if there was a better forum for presenting findings to a broad spectrum of program specialists who could utilize learnings from existing programs as they plan and implement future programs.

l. There is a classic difficulty that the different disciplines have in finding a common language and in collaborating across boundaries. "Turfism" is apparent in the lack of willingness on the part of organizations to share information with potential competitors or to share evaluation findings that reflect unmet expectations or undesired outcomes.

m. Delivery of youth development programs at the local level is not standardized. It is difficult to identify nationally uniform outcomes to assess given the diversity in goals and outcomes in local youth development programs.

n. Unreasonable or unclear requirements from funders lead to fear of loss of funds and often make careful evaluation seem risky. Funders often expect outcome evaluation to be conducted with limited time and limited money.

o. Invidious comparisons exist between "real" or "hard" science (i.e., biomedical) and "soft science" (i.e., social and behavioral).

p. The success of a program may not be immediate, but may manifest itself many years later (as shown by the long-term follow-up of Headstart).

q. The "null hypothesis," effect/no effect approach to evaluating youth development is still prevalent. Evaluators attempt to discover if the program had a measurable impact that did not occur by chance. This approach, effective in disciplines where outcomes are discrete and immediate, is less effective in youth development where programs are likely to have multiple, overlapping and indeterminate outcomes. When rigorous evaluations demonstrate "no statistically significant effect," evaluators are likely to fault the program rather than their own methods.

r. Although many evaluators, funders, and program directors are recommending greater use of qualitative evaluation methods, these studies have their own set of problems:
   - o weak legitimacy in the eyes of traditional quantitative researchers.
   - o concerns about what can be generalized from such studies.
   - o labor-intensiveness of methods.
   - o tremendous variation in competence of evaluators' use of qualitative methods.
   - o indiscriminate use of vivid case examples to prove

points rather than well-grounded conclusions based on systematic analysis.

- lack of knowledge on the part of funders and policy-makers--they don't know what to look for, or to demand, in a good qualitative study.

s. Conceptual and methodological problems occur when multiple programs are implemented in one community and these programs serve a common group of young people. Often youth who participate in a program being evaluated for one organization also participate in programs offered by other organizations. The evaluator is charged with determining whether any observable effect on the young person is due to the specific program being studied.

t. There is a lack of culturally-appropriate intervention strategies by service providers and evaluators.

u. Too little attention is paid to involving youth in the evaluation process.

v. There is a tendency among researchers and critics to get weighed down in debates about methodological soundness and techniques.

3. How can we build on and strengthen program evaluation efforts in the next decade?

a. Identify more fully the range of possible effects that programs may have on young people. Seek answers to the question: "What kinds of effects can programs of different kinds have on young people who have different internal traits and different external circumstances in which they are living?" Spend more time talking to program directors, staff and participants to find out what impacts they believe programs are having on the lives of youth.

b. Because of the tremendous variety in youth development programs, it would be useful to understand which program characteristics or dimensions (intensive vs. intermittent participation, program content and/or philosophy, staff-youth mode of interaction, etc.) are associated with particular outcomes among participants. It is quite possible that the mechanisms through which many youth organizations achieve positive results are at the global level--program philosophy, structure, staffing and practices.

c. Define the outcomes (with an emphasis on attitudes and behaviors we want to promote in addition to those we want to eliminate or postpone) and identify and label uniformly the inputs we think quality youth development programs offer that can be measured in doses. In order to do this we must define the essential elements or "best practices" of a quality youth development program. Eventually we need to understand if best practices vary for programs of different kinds or for different youth.

d. Given the diversity in goals and outcomes for youth development programs at the local level, conduct hundreds of local evaluations and seek to identify overall trends by meta-analyses of evaluation findings.

e. Provide a more sophisticated linkage between program processes and outcomes. Identify the contextual and program factors that make for success and failure. Conduct good causal studies based on qualitative data. We need to move beyond the black box type of evaluations that tell us a program is effective or ineffective but do not tell us why or why not.

f. Identify any negative influences that programs might have on young people. For example, what are the possible consequences of programs that are poorly conceived, structured, or implemented?

g. Conduct deeper studies of actual program implementation: what programs actually look like at the point of delivery, including close-up portraits of adolescents in their life contexts. Perhaps conduct a careful documentation of "a day in the life" of a youth program.

h. Do a better job of linking quantitative and qualitative data bases drawn from the same programs.

i. Link ethnographic research with experimental and quasi-experimental research by promoting multi-disciplinary research teams.

j. Broaden program evaluations to include a community context.

k. Design coherent, well-grounded case studies and strengthen multi-case study analysis methods.

l. Reduce the distance between program realities and larger-scale planning. Use findings from existing studies to make recommendations for public policy and programmatic decision-making.

m. Develop a wider range of formative evaluation functions which would: 1) heighten local understanding of program functioning, 2) enhance local self assessment and improvement efforts, 3) provide data-linked technical assistance and in-service training.

n. Increase visibility and credibility of program evaluation within the field of youth development and also in the educational and social service fields.

o. Program administrators and funders must insist that evaluations produce useful results that are likely to advance the state of the art. Rather than choosing evaluators on the basis of their credentials or their use of venerated evaluation methods, youth development agencies should find consultants who will conduct evaluations that actually help program staff and participating youth.

p. Recognize that working with youth is a long-term investment. Therefore, the impact of youth organizations can best be measured in the long term.

q. Hold regularly scheduled meetings of evaluators, and perhaps develop a clearinghouse of evaluation reports and measures.

r. Develop evaluation methods that take into account the reinforcement produced by participation in multiple programs (conducted by different youth development organizations) rather than treating the coexistence of programs as a source of study "contamination."

s. Significant progress will require collaboration of the major youth development organizations and the major funders of youth development programs to reduce the risks and increase the capability of improving program evaluation.

EVIDENCE OF PROGRAM SUCCESS
AMONG NATIONAL YOUTH ORGANIZATIONS

The following summary is drawn from the national organization profiles and interviews that are part of the work of the Task Force on Youth Development and Community Programs. These summaries are responses to the profile question: Do these organizations have any evidence of their program's success?

American Camping Association

ACA leadership cite four factors as evidence of the success of the traditional camping experience: (1) the longevity of the industry; (2) the breadth and diversity of the experiences offered by ACA members; (3) anecdotal evidence, including personal stories from large numbers of campers about the effect of their participation; (4) theoretical evidence, such as the way camps generally employ developmentally-appropriate and active learning modes (ACA staff cite the Center for Early Adolescence research, in particular).

ASPIRA

In 1989 ASPIRA staff initiated a follow up survey to measure the effect of their Public Policy Leadership Program and found the following:

94% of participants were enrolled in school (60% in college, 40% in high school);
32% were involved in school government;
63% were involved in a school club;
More than half held offices in groups in which they were involved;
75% reported being more assertive and self-confident as well as developing leadership skills after involvement in the program.

Big Brothers/Big Sisters

In 1990 Big Brothers/Big Sisters of America contracted with Public/Private Ventures, a Philadelphia-based program development and technical assistance organization, to conduct an evaluation of its basic program.

The evaluation, which will be implemented over a four year period in 15 sites, involves four separate but interrelated studies: (1) a study of the effects of the relationship with a Big Brother or Big Sister on the lives of participants; (2) a study of volunteer recruitment and screening; (3) a study of the administrative and operational practices that comprise the BB/BSA program model; and (4) a qualitative examination of the interactions that take place between the adults and the youth with whom they are paired.

Public/Private Ventures staff have spent considerable time getting to know the culture, programmatic methods, and needs of Big Brothers/Big Sisters of America, and from the perspectives of both organizations, this has been time well spent.

This evaluation has potentially great significance because the effort is designed to assess the effectiveness of the organization's "core service," whereas most evaluations are able to evaluate only one component of an agency's total program.

Boy Scouts

BSA has no outcome evaluation of its programs. Many famous alumni of Boy Scouts have supported the organization's recruitment campaigns by putting in a good word about their Boy Scout involvement as youth. Boy Scouts of American also cites dissemination and utilization of its materials as a sign of success.

Boys and Girls Clubs of America

Boys and Girls Clubs of America contracted with Steven Schinke of Columbia University to conduct an evaluation of the implementation of its "SMART Moves" substance abuse prevention program in public housing projects. The evaluation design compared the impact of three situations: a housing project with no Boys and Girls Club; a housing project with a Boys and Girls Club but without "SMART Moves"; and a housing project with a Boys and Girls Club that offered "SMART Moves."

The evaluators found that, while the differences in impact between the Clubs without "SMART Moves" and Clubs with "SMART Moves" were not great, there were substantial differences between the housing projects that had Clubs and those that did not in relation to positive outcomes for youth, for parents, and for the surrounding community. "For youth who live in public housing and who have access to a Boys and Girls Club, the influence of Boys and Girls Clubs is manifest in their involvement in healthy and constructive educational, social and recreational activities. Relative to their counterparts who do not have access to a Club, these youth are less involved in unhealthy, deviant and dangerous activities," noted the evaluators. Data from the evaluation showed that adult residents of public housing were also beneficially affected: compared with parents in the control (no-Club) sites, adult family members in communities with Boys and Girls Clubs were more involved in youth-oriented activities and school programs. For adults and youth alike, Boys and Girls Clubs appeared to be associated with an overall reduction in alcohol and other drug use, drug trafficking, and other drug-related crime.

In 1986 Boys Clubs contracted with Louis Harris to do a survey of Boys Clubs' alumni. This survey is used by the Boys and Girls Clubs of America as evidence that participation in Boys and Girls Clubs has positive effects.

The results of the Louis Harris survey were reported in the Summer of 1986. According to the survey, nine of ten alumni reported that Boys Clubs involvement had a positive effect on their lives, gave them skills for leadership, helped them get along with others, and influenced their success later in life.

Alumni remembered the Club as one of the few places in their neighborhoods where they could go to participate in organized activities and find refuge from the street. Three out of four alumni said Clubs helped them stay out of trouble with the law. Seven out of ten said their involvement with Clubs helped them avoid problems with drug or alcohol abuse. Two out of three former members interviewed are now professionals, managers, proprietors, or skilled workers.

Camp Fire Boys and Girls

Camp Fire has done no outcome evaluation of its programs. However, they do conduct extensive field-testing of new program materials before they are published and disseminated.

COSSMHO

Some evaluation is built into each national COSSMHO program. Its most rigorous evaluation to date was a process and outcome evaluation of the OAPP-funded Strengthening Families program. All three demonstration sites (local agencies in Kansas City, MO; Mission, Texas; and Puerto Rico) participated in the outcome evaluation, which employed an experimental design and showed immediate positive results, including more effective communication within families about sexuality and pregnancy. These behavioral gains were not sustained at the three-month follow-up, however. The organization is planning to conduct a rigorous outcome evaluation of its new inhalant abuse prevention program.

Child Welfare League of America

CWLA does not have a "program" per se, but it fosters and publicizes research on child welfare service effectiveness as part of its ongoing work.

Girls Incorporated

Girls Inc. has conducted rigorous outcome evaluations of its Friendly PEERsuasion and Preventing Adolescent Pregnancy programs; both evaluations have shown positive results.

The Friendly PEERsuasion program was evaluated by staff of Abt Associates. The evaluation indicated that: (1) the program significantly reduced the incidence of drinking among enrollees and the onset of drinking among enrollees who had not previously drunk alcohol; (2) the program led enrollees to disengage from peers who smoked or took drugs; (3) the program may have been more effective with younger enrollees in either reducing substance abuse or reducing associations with substance abusing peers; however, the only significant difference in effect was on the combined incidence of any substance abuse.

The evaluation of the Preventing Adolescent Pregnancy Program was conducted by Girls Incorporated's own evaluation staff (from its National Resource Center). A three-year longitudinal study of the outcomes of participation in the four-part program indicated that: (1) Girls who participated in "Growing Together" (parent-daughter communication workshops) were only half as likely to have sexual intercourse for the first time as girls who did not participate; (2) Girls who participated in the entire "Will Power/Won't Power" program were only half as likely to have sexual intercourse for the first time as girls who did not participate; (3) "Health Bridge" participants reported having sex without birth control only about one-third as often as nonparticipants; (4) Young women who participated in the entire "Taking Care of Business" program were only half as likely as nonparticipants to have sex without using birth control. They were also less likely to become pregnant than nonparticipants.

Girl Scouts of the U.S.A.

Girl Scouts has conducted no outcome evaluation of its programs, although they conduct extensive field-testing of program materials before they are published and disseminated.

In 1990, a study of a nationally representative sample of Junior, Cadette, and Senior Girl Scouts conducted by Louis Harris and Associates revealed interesting data to support program success.

The organization has its most powerful impact on minority girls, especially blacks. A full 89% of black current and former Girl Scouts report that the organization is at least somewhat important to them, compared with 86% of Girl Scouts who are Hispanic and 79% who are white.

By Girl Scout participants' own account, they do better academically than other girls.

Over 75% of former and current Girl Scouts surveyed said that Girl Scouts taught them about things like good health and safety; helped them do something good for their community; taught them to have more respect for other people; and helped them to gain new skills.

Over half of respondents said Girl Scouts made them more sensitive to the needs of other people; helped them learn the difference between right and wrong; helped them feel better about themselves; helped them meet girls of different races, ethnic, or cultural backgrounds from their own; and offered them a group of girls they could turn to when they need help and advice.

Sixty-six percent of the Girl Scouts surveyed are "very satisfied" and another 22 percent are "somewhat satisfied" with the adults who work with their troop. Girl Scouts who are Black were most satisfied, with 84 percent saying "very satisfied."

Elements most valued by current and former members were fun and friendship.

Older Girl Scouts, especially Senior but also Cadette Girl Scouts, see considerable opportunities for decision-making, involvement and leadership in their troop meetings; they see these opportunities as much less available in their school classroom. For example, 74 percent of Senior Girl Scouts feel their troop leaders very often listen to what they say while only 42 percent feel their classroom teachers very often listen. Similarly, 66 percent of Senior Girl Scouts say they very often make decisions about what goes on in their troop but only 25 percent often make decisions about what goes on in their classroom.

Compared to girls in a national sample (Girl Scouts Survey on the Beliefs and Moral Values of America's Children), the Girl Scout sample was more likely to make sound moral choices in hypothetical moral situations. Although a one-time cross-sectional study cannot imply causation, it is evidence that merits further study to see if Girl Scouting does influence moral decision-making. The study did find that girls who most believed in the usefulness of the Girl Scout Promise and Law in their lives were much more likely to make good moral choices.

4-H

Because of its tie to Land Grant Colleges and Universities, 4-H has access to the research and evaluation capacity of those institutions. Many doctoral dissertations have evaluated the impact of participation in various aspects of 4-H programs, including long-term impact. The Youth Development Information Center in Greenbelt, MD has a close-to-complete collection of these dissertations. Because there is really no national 4-H program, it is difficult to apply these findings to the entire system. However, these studies provide general support for the value of participation in the 4-H program, and they are often cited as support for participation in other kinds of non-formal education.

The University of Kentucky performed a survey of state farmers and found that those who had been 4-H members had higher educations, higher farm sales, and higher farm incomes than non-participants. They were more likely to use innovative farming techniques. Twenty-seven percent of Kentucky's farmers had been 4-H members, averaging a 4-1/2 years membership, and 9 out of 10 surveyed rated the 4-H experience "worthwhile."

The Extension Service of the USDA funded a national study of alumni in 1987. Phone interviews were conducted of a random sample of 710 farmers who had been 4-H members, 743 farmers who had been members of other organizations, and 309 farmers who had never participated in youth organizations.

One finding was that 4-H alumni and alumni of other youth organizations were more alike than non-participants in race, age, family income, and number of children participating in youth organizations. Non-participants tended to be minorities with lower family income and less education.

A second finding was that 53% of 4-H alumni belonged to other organizations as youth as well (39.6% belonged to their church youth groups). Among these alumni, 4-H rated slightly higher than other organizations in developing knowledge and skills and imparting feelings of self-worth. Other organizations rated higher than 4-H in developing leadership skills.

Average age of participants when they joined 4-H was 10.6 as compared to 9.5 for other organizations. Participants on average were members of 4-H for four years, as compared to six years for other organizations. Those who joined the earliest stayed in the longest. 57.9% respondents said they did not join 4-H because it was not available.

Participants in 4-H rated contact with other people as the most useful experience they received. Of the 59% of alumni who dropped out of 4-H, 44.4% did so because 4-H did not meet their interests and 21.3% thought the program was for younger kids.

Participants in youth organizations were more likely to be involved in community activities as adults than were non-participants (no difference between 4-H alumni and other youth group alumni).

Conclusions of the study were that three factors could improve the growth and impact of 4-H: (1) to enhance the visibility of 4-H ; (2) to design programs for older youth; (3) to offer more opportunities to develop leadership.

Junior Achievement

All Junior Achievement programs are independently evaluated and updated on a three-year cycle. These evaluations consist of administering questionnaires to samples of teachers, consultants, and students to assess the effectiveness of the programs.

A 1990 study of Junior Achievement's Project Business (PB), a course that supplements 8th grade social studies classes, was conducted in Chicago schools by an independent contractor. The study evaluated the content, activities and support of the PB curriculum in order to determine its effectiveness and appropriateness in Chicago schools. The study found that the PB program was well received--95% of teachers and consultants considered it a rewarding experience. The activity and discussion-based learning approach involved in the course overcame any obstacles of working with young people who were not good readers. Sixty-three percent of students sampled said they learned a lot about starting their own business. No systematic weaknesses were found in the manual or curriculum, but it was suggested that more effort was needed to produce materials that reflect the students' environment. Similar findings were reported in a national study of the Project Business course conducted in 1988. In this study it was found that over 75% of respondents enjoyed the course and over 85% of teachers and business consultants would teach the course again. Teachers in low-income urban classrooms were more likely to perceive the course as being an effective teaching resource.

In 1989 Junior Achievement hired an independent contractor to evaluate their Applied Economics (AE) course, a semester long course in economics for high school students. The study found that teachers, consultants, and students had a strong, positive overall perception of the AE experience. Teachers considered AE to be more interesting and worthwhile to teach than other high school economics courses. They considered the outside consultants to be an essential component of the course. Eighty percent of teachers and consultants would recommend the course to their peers. Ninety-three percent of teachers and eighty-two percent of consultants sampled would teach the course again. Findings show that teachers and consultants were only moderately satisfied with the training they received to teach the course. In student assessments of the course, 69% said their experience was excellent or good, and 37% said it was better than other courses (18% said it was worse).

Findings of a 1987 study of the JA Business Basics course, designed to introduce 4th, 5th, and 6th graders to basic principles of economics and business, show that overall, people involved in the program thought highly of the experience. The course could be improved in several areas, based on teachers' assessments, and the consultants' training could be improved.

NAACP

No real outcome evaluations have been conducted of any of the NAACP's programs. The organization's literature cites two kinds of evidence for the success of its efforts: (1) regarding its principal focus, promotion and protection of civil rights, it cites its impressive and long-standing record of success in initiating and winning course cases (such as Brown vs. The Board of Education, 1954); (2) regarding the leadership development emphasis of its Youth and College Division, the NAACP cites its list of famous alumni, including Julian Bond, Roy Wilkins, Vernan Jordan, Ralph Bunche, Andrew Young, and Thurgood Marshall.

National Network of Runaway and Youth Services

One goal of most runaway shelters is to reunite runaway youth with families. Measured against this goal, their work is quite successful. A 1987 study revealed that 53% of youth served by federally-funded runaway shelters returned home and that another 32% were "placed positively and appropriately" in foster or group homes, independent living centers or other type of treatment programs. Less than 10% of youth served returned to the streets.

National Urban League

The Urban League regularly conducts both process and outcome evaluations of its programs. For example, its National Education Initiative, now in its second five-year phase, was evaluated by a third party (Dr. Cardwell) at the end of its first five-year phase, in July of 1990. Because of NEI's orientation toward systemic change, this evaluation measured changes in dropout rates among African American students, as well as other outcomes. NUL has outcome data concluding that 89% of the people who received direct training in the organization's employment programs were placed in jobs. NUL currently has no outcome date on any of its national youth programs since these initiatives are all new, but it does expect to conduct such evaluations in the future.

Salvation Army

Some of the Salvation Army's nationally-developed programs have been the subject of outcome evaluations. For example, the Bridging the Gap (life skills training) program was evaluated by an outside evaluation team from the Center for Informative

Evaluation, which conducted both a process and an outcome evaluation. The research showed that participants (650 teenagers in 23 sites) made substantial gains in knowledge about themselves, the community and its resources.

WAVE, Inc.

Throughout the history of WAVE in Communities, they have constantly evaluated the program--measuring job placements, GED attainments, and other positive outcomes.

WAVE In Schools was evaluated by third party evaluators (Institute for Educational Leadership) during its demonstration year. First year results included "positive changes in attitudes, behaviors, and academic achievement for the majority of students." Results also indicated that students improved their reading and math levels, as well as their scores on instruments aimed at self-esteem, pre-employment and work maturity. Absences from school and suspensions also decreased.

YMCA

The YMCA has conducted no outcome evaluation of its youth programs. However, the organization does have a research director on its national staff.

YWCA

The YWCA has conducted no outcome evaluation of its youth programs.

SUMMARIES OF SELECTED ARTICLES

SUMMARY OF "EVALUATING SOCIAL PROGRAMS: WHAT'S BEEN LEARNED"
By Milbrey Wallin McLaughlin
(Unpublished Paper Prepared for the Ford Foundation, Fall 1987)

Models for evaluating social programs have evolved in little or more than twenty years from an initial and single concentration on testing and measurement of student achievement to the current interest in providing information to support policy and program decision-making. In the middle third of this century, Ralph Tyler, the "father of educational evaluation," changed the field with his conception of evaluation as comparison between intended and observed program objectives.

In the mid 1960s Great Society initiatives mandated an unprecedented amount of evaluation activity. Ambitious programs sponsored by the Elementary and Secondary Education Act (ESEA), as well as smaller federal initiatives, such as Head Start, Follow Through and Right-to-Read, all required evaluation of local efforts.

Federal-level insistence on evaluation was thrust upon a largely unprepared field. Initial evaluations of federal education programs built on Tyler and traditional social science models that had been developed in academe to assess the outcomes of clinical experiments or laboratory trials. These evaluations typically used experimental methodology that paid little attention to the context of program activities or the processes by which program plans were translated into practice.

Early approaches to the evaluation of social programs assumed that:

- randomization and the use of control groups were the sine qua non of "good" evaluation.
- there was independence of observed effects, a relatively static program environment and stable program outcomes--presumptions rooted in a clinical model.
- the substantive model of practice was known.
- a direct relationship existed between treatment (or program inputs) and effects (or program outputs).
- the "black-box" of program setting contained few powerful program effects.
- program activities were a discrete aspect of their institutional setting that could be studied in isolation.

Lessons from the Field

The initial spate of program evaluation generally reached discouraging conclusions of "no significant differences," but it was impossible to understand whether the absence of measurable effects was the consequence of poor program design or attributable to the evaluation model. These evaluations provided scant information about how programs were put into practice, or the effects masked by summary statistics aggregated at the program or school level. Unable to find much use for these results, both practitioners and policymakers generally ignored these early evaluation efforts. The evaluation community also agreed that these assessment efforts failed to get at what mattered to program outcomes--local choices, conditions and practices--and that they were not conducted in ways that were useful to policy and practice.

The first round of educational evaluation, in short, provided little information about why programs failed or succeeded, what promising strategies might be, or how successful efforts could be carried out in other settings.

Theories of Action Reconsidered

Experience generated by these early evaluation efforts led to the following lessons about how change occurs in organizations:

- **Implementation dominates outcomes.** Local choices about how to put a policy into practice determine the extent to which a policy or program fulfills its promise, whether the benefits reach the intended target group, or in fact whether a new program is carried out as planners intended, or even at all. Further, local factors often beyond the reach of policy (available resources, capacity or motivation, for example) shape these choices in fundamental ways.

- **Implementation is a multi-stage, developmental process.** Program or policy implementation proceeds though analytically distinct stages involving different actors, different issues and different consequences. Implementation is a complex process of institutional and individual learning. In most cases, institutions need to learn the rules of the game before substantial and confident attention can be devoted to learning about how to make practice more effective. Effective implementation of new policies and practices takes time.

- **Social programs operate in a fluid context.** Social programs function in a dynamic and often unpredictable environment that changes both the nature of the problems addresses by programs as well as the resources available to address them. Within this context, neither success nor failure are fixed, nor are the resources that shape program outcomes.

- **There are few "slam-bang" effects.** Change often is marginal and incremental. The short-term significant differences are diluted as the implementing system responds to changed practices and adapts to new routines. Significant program-related changes may appear over time, but seldom do program effects appear in the one or two year time horizon adopted by many program evaluations.

- **In practice, social policies and programs have multiple goals.** Programs have political and bureaucratic consequences in addition to the service goals that are typically the sole focus of traditional evaluation models. Each of these goals is likely to be assessed differently by different actors in the implementing system.

- **Social problems are complex and not well-understood.** Simplistic conceptions of social problems to be solved led to single-focus policies such as compensatory education, Right-to-Read Programs and bilingual education. Social problems addressed by public policy have multiple, complex roots and definitions of underlying problems many times are not clear.

Evaluation Reconsidered

Conceptions of program evaluation have moved away from the laboratory model toward a more global, dynamic, decision-oriented approach:

- **The unit of analysis is the implementing system.** The system responsible for carrying out and supporting social programs consists of interrelated components. Evaluation of one component needs to assess how it fits with others. An understanding of the contextual factors and relationships can be critical for interpreting project outcomes.

- **Evaluators need to cast a wide net around a project,** especially in the early stages of its operation, in order to capture important main- and lower-order influences. This broader view also captures unanticipated consequences or effects which are associated with program activities and can have major import for evaluation conclusions.

  There is a need to integrate micro and macro levels of analysis. The latter focuses on the larger systems view and raises questions of program implementation; the former involves program theory about treatment or program activities and their consequences for participants. This integrated perspective is essential in order to distinguish between failure of theory and failure of implementation, and to understand the conditions under which project activities occur.

- **There is no one best model.** Evaluation activities and objectives need to fit program realities and the context of decision-making.

- **Evaluation models must fit the stage of program operations.** Social programs proceed through stages of adoption, implementation, assessment, and institutionalization. The issues for program operations and so for evaluation differ at each stage. In particular, premature questions of program effects are both unwarranted and potentially destructive.

- **Evaluation designs must focus on contingent aspects of program operations and outcomes.** Program or project evaluation, consequently, needs to examine and elaborate the conditions under which observed activities and outcomes occurred and to assess the importance of aspects of the institutional setting to components of the program.

- **Use multiple evaluation methods.** Debate about the virtue or value of quantitative methods has moved from "right" or "wrong" to when and how. Qualitative methods are suited to collecting rich information on the processes of program implementation; quantitative procedures can generate a standardized assessment of project outcomes both within and among project sites.

Contemporary Approaches to Evaluation

The current challenge among evaluators of social programs is to select the evaluation approach most appropriate to a given program or decision setting at a given point in time. Choices about which evaluation design to use depend on at least four broad considerations: 1) the evaluation purpose; 2) the decision context; 3) the stage of program development; and 4) the status of the program theory or knowledge base underlying program activities.

- **There are multiple purposes for undertaking an evaluation --** policy formulation, program implementation, accountability and program improvement or policy revision. Each calls for a somewhat different evaluation approach.

- **The decision context of evaluation also shapes choices about design.** Is the primary client a practitioner who needs information about implementation costs for an upcoming board meeting or a legislator who needs to understand the benefits of a program as well as the broad costs in political and institutional terms?

- **Design choices are contingent upon the stage of program development.** The field has learned that "impact" evaluations are inappropriate until sufficient time has elapsed for a program to be implemented, achieve some measure of stability, and operate in a manner that program theory assumed would generate expected outcomes.

- **Finally, the status of program theory or the knowledge base underlying program operations has consequences for choices about evaluation.** An experimental or quasi-experimental model, for example, assumes a relatively well-developed theoretical base. The purpose of evaluation, in this instance, is to examine the theoretical expectations in practice.

Contemporary evaluators have a diverse assortment of designs or evaluation approaches from which to choose, given purpose, context, program stage and strength of theory. These diverse approaches to evaluation differ on many dimensions. Chief among them are instrumentation (from highly standardized, closed evaluation instruments to open-ended, ethnographic inquiry), role of the evaluator (from educator to management consultant to assessor to advocate), role of client (from active stakeholder and collaborator to passive recipient of evaluation product), to overall design (from experimental or quasi-experimental to exploratory), focus (on process--formative evaluation--or outcome--summative evaluation). Each of these dimensions corresponds to the contingencies upon which evaluation choices are based--purpose, decision-context, stage of program development and status of theory or knowledge base.

Promoting Use

Learning about how evaluation-based information is used has led to revised notions of "useful knowledge" and strategies for enhancing utilization. Evaluators have become conscious of the so-called "two cultures" problem, or the issue of correspondence between the conceptual world of the evaluator and the practical world of the policymaker or practitioner.

In the early years of program evaluation, results arrived too late to be of use. One important lesson involved tying evaluation reporting tightly to policy or decision timelines. Closely associated with timeliness is the extent to which evaluation addresses the specific needs of decision-makers.

Use also includes influence on how policymakers or practitioners think about a problem, impact on the general climate of opinion surrounding a policy issue, or persuasion--using evaluation to provide support for a particular position or program.

Where evaluations are intended to inform policy or practice in the short term, utilization can be enhanced by moving toward a collaborative model in which evaluator and client work together to identify central questions and fundamental assessment criteria, clarify definitions and concepts, and establish a format and schedule for reporting.

Collaboration can have implications for the evaluation process leading to modifications in the design as issues are clarified and assumptions tested in practice. From the evaluator's perspective, this collaborative model should also involve preparing the client to use evaluation results.

Evaluating Programs for At-Risk Youth

The realities and issues that complicate the evaluation of social programs in general raise especially difficult concerns for evaluation of programs for at-risk youth. Briefly:

- There are few agreed-upon definitions of "at-risk youth."
- Problems of youth at-risk result from complex, interrelated, multiple conditions--family patterns, changed economic realities, disintegrating social institutions, constrained social services all contribute to the problems. What then is the most appropriate knowledge base or theoretical tradition to use in developing program strategies? Can a single program respond to the complex pathology responsible for risk?
- Both problems and programs generally have been defined by actors in the mainstream, not by members of the target population or direct stakeholders. Risks of misunderstanding the problem and so misspecifying solutions are considerable.
- The constituency for serving at-risk youth is uncertain and politically ineffective.
- It is likely that practices effective for at-risk youth will be unconventional. Promising practices may thus face rejection by the social system or institutional setting best situated to implement them.

Efforts to evaluate programs for at-risk youth, in short, amplify all of the problems evaluators confronted in efforts to evaluate social action programs in the past and add others associated with the non-mainstream character of the problem and likely solutions. Evaluation responses, accordingly, must be especially creative, thoughtful and eclectic.

SUMMARY OF
"ISSUES IN EVALUATION FOR DISCUSSION AT THE SECOND INTERNATIONAL CONFERENCE ON URBAN SAFETY, DRUGS AND CRIME PREVENTION"

Paris, November 18-20, 1991
The Milton S. Eisenhower Foundation

The purpose of this document is to summarize the Eisenhower Foundation's assessment of evaluation methods. The objective here is to review the successes and failures of the evaluation methods used to assess the local programs.

Community Surveys

One method used extensively -- the community victimization surveys in both test and companion neighborhoods -- often proved too expensive for what they told us.

Our researchers polled neighborhood residents in depth before and after the program to find out about crime, fear and many other issues.

The Foundation was, on balance, reminded of the observation of Dr. David Hamburg, President of the Carnegie Corporation, that evaluation can divert too many scarce financial resources away from actual program strategies.

While community surveys of citizens measured community change, they were inadequate for measurement of change over time among specific individual high-risk youth in the programs.

Feedback From Street Level Program Directors

Most program directors agreed that the community survey supplied information that was useful for planning their initiatives. They also indicated that the process information on day-to-day implementation lessons, such as the importance of technical assistance, was very useful.

At the same time, most program directors said they had wanted to participate more in the design of the surveys and evaluations of them to ensure that their own definitions of success were taken into account. Program directors called for evaluators to play less the role of "experts" and more the role of "collaborators" in the future.

Some program directors said there was a "negative response" by neighborhood residents to the content and style of administering the community surveys which led some of the residents to respond untruthfully. Local directors also felt the surveys should have been conducted with same-race interviewers.

Not uncommonly, program directors believed that the evaluation did not adequately look for relationships between crime and fear, on the one hand and neighborhood-wide progress, like housing rehabilitation, on the other. In this vein, program directors asked that crime-related measures be viewed not only as outcomes but also as symptoms of more deep-seated community problems, like unemployment. There was also a need to extend evaluations over longer periods of time than thirty months to uncover impacts.

Measures of Individual Change

To move beyond the limitations of community victimization surveys, studies of changes over time among individual program youth were undertaken -- by Rutgers University. Program directors found these Rutgers case study evaluations to be sufficiently tailored to their street realities.

Future Directions

The Eisenhower Foundation intends to follow programs over longer periods of time than the thirty-month planning and implementation period which we originally assumed was the minimum to observe effects. We will, among other refinements, take more measures of program youth over longer periods of time (a "repeated measures" design).

Specifically, future Eisenhower evaluations will seek to cover thirty-six to forty-eight months of planning and implementation, incorporate both process and comparison group impact measures, and trace both change among individual program youth and change in the community where the program is located.

Since the Rutgers case study format was relatively inexpensive, it is of critical importance for the future. In all new programs, the Eisenhower Foundation will seek a balance--lower cost evaluations with findings that remain valid and reliable. The Foundation will also follow the advice of Professor Donald Campbell, dean of program evaluators in the U.S., who asks for common sense assessments that integrate the views of both outside evaluators and committed practitioners.

Excerpt from Schorr, Lisbeth. (1988). Within Our Reach: Breaking the Cycle of Disadvantage (pp. 267-270). New York: Doubleday.

Knowing What Works

Because of the widely held view that we really know very little about such matters as crime, teenage pregnancy, and school drop-outs, reliable evidence about interventions that work has become more important than ever. Twenty years ago, when social policy was being formulated in an atmosphere of boundless optimism, the combination of a little theoretical research, fragments of experience, and a lot of faith and dedication was enough to justify a new social program. Today budget deficits, fears of wasting money and perpetuating dependency, and a gloomy sense of social problems beyond solution have combined to reinforce the demands of the keepers of the purse strings to see tangible evidence of effectiveness as a condition for support of any social program.

Unfortunately, the reasonable demand for evidence that something good is happening as a result of the investment of funds often exerts unreasonable pressures to convert both program input and outcomes into whatever can be readily measured. This rush to quantify, which engages funders, policymakers, academics, policy analysts, and program administrators alike, has had damaging effects on the development of sound interventions aimed at long-term outcomes. Programs are driven into building successes by ducking hard cases. Agencies shy away from high-risk youngsters, who provide scant payoff for effort expended when it comes to bottom-line totals. Energy is diverted into evaluation research that asks trivial questions and sacrifices significance to precision.

Pressures to quantify have crippling effects on the development of the kind of programs most likely to help high-risk families. Current methods of demonstrating effectiveness do not capture the essential extra dimension that characterizes successful programs. Organizations are pressed to shape their objectives and methods of intervening with an eye to easy measurement, and cannot be blamed for choosing to narrow rather than broaden their efforts.

Many of the most effective interventions with high-risk families are inherently unstandardized and idiosyncratic. Many agencies have found a mix of services, adaptable to different sites and responsive to particular family needs, to be an essential component of effective interventions. When a home visitor, for example, responds flexibly to a family's unique problems, the unique outcome may be just what the family needs but what the evaluator dreads. (The young mother worried about the illness of a grandparent seeks the advice of the home visitor, who responds to this concern instead of teaching the mother how to read a thermometer, as planned. Will the evaluator be able to capture the young woman's greater comfort and trust, and their consequences for mother and child?)

Educators working with disadvantaged children find that "these kids can't learn until they learn to trust" and that "sustained intellectual growth depends on the quality of relationships established between parent, teacher, and child." Are program objectives like the acquisition of trust or the development of warm personal relationships, found to be essential attributes of virtually all programs serving high-risk families, to be sacrificed because they are so much harder to reduce to quantifiable terms than is performance on multiple-choice or IQ tests?

Some program outcomes, such as the effect of preschool education on increasing the chances of high school completion or effect of family support on reducing the incidence of delinquency, are difficult and expensive to document because of the distance in time and place between intervention and outcome. Are the interventions whose payoff is difficult to document to receive less support as a result?

For many services, how they are delivered is as important as that they are delivered. For example, it has been conclusively established that responding to patients' and families' psychological needs has favorable effects on health outcomes, and that the physician's "skillful listening, empathy, warmth and attentive interest" are central to good and appropriate child health care. Yet the subtle "how" eludes us. Policy-making tends to remain on the more solid ground of numbers of children covered by health insurance and numbers who see physicians, even though these numbers tell us little about the adequacy of the health care they receive. Pushed to rely on what is countable, we have come to regard access to health services as an adequate measure of effectiveness. Similarly, the number of dollars spent on education, child care, or social services becomes a proxy measure which is quickly equated with effectiveness--because it is often the only window on what is actually happening.

The rush to quantitative judgment, with its demands for immediate results, also interferes with orderly progress in developing complex programs. Professor Donald Campbell, considered by many the dean of program evaluation, says a new norm is needed to replace the current practice of prematurely evaluating programs not yet working as their staffs intended. The principle he proposes is "Evaluate no program until it is proud." By not insisting on formal evaluations until program personnel have themselves concluded that there is "something special that we know works here and we think others ought to borrow," Dr. Campbell believes the sum total of useful "borrowable" information would be vastly increased.

Professor Campbell also endorses an approach to the use of information that forms a fundamental premise of this book: judgments and decisions should be based on a thoughtful appraisal of the many kinds of evidence available. That means relying not only on quantitative but also on qualitative information, not only on evaluations by "objective" outsiders but on the experiences of committed practitioners, not on isolated discoveries but on understanding how consistent the findings are with other knowledge. Relying on common sense, prudence, and understanding in interpreting evidence does not mean sacrificing rigor in assessing information. But applying human intelligence may bring us closer to policy-relevant conclusions than reliance on numbers that have been manipulated in ways that ultimately conceal a basic ignorance of what is really going on.

SUMMARY OF THE APPLICATION OF PROGRAM EVALUATION IN THE MANAGEMENT OF NON-PROFIT SECTOR: AN EXPLORATORY STUDY

Russy D. Sumariwalla
Martha E. Taylor
United Way of America

March 1991

ABSTRACT

National non-profits were asked to provide information on the evaluation activities of their members/affiliates or grantees. The most frequent types of evaluation are measurement of volume of program delivery and compliance with standards. Least frequent are assessment of participant satisfaction and assessment of program outcomes or results. The two greatest barriers to conducting assessment of program outcomes/results were perceived to be lack of financial support and lack of staff with necessary skills. The best strategy to increase such evaluation was increased support by funders. Other strategies frequently mentioned were increased training, research on measures/instruments and access to organizations with expertise. This study also identifies variations in types of evaluation conducted, staff support and perceptions of needs among different types of non-profits.

I. BACKGROUND

At United Way of American the quest for evaluation tools can be traced back to early seventies with the launching of "The House of Accountability." In fact, effectiveness assessment tools were widely advertised as the "roof" on "the House" implying the completion of a series of tools for the use of local United Ways and other human service organizations. Other elements of the "House of Accountability" consisted of: Service Identification, Definition, and Classification; Accounting and Budgeting Guides; Campaign and Allocation Analyses; and Needs Assessment.

Early efforts at United Way of America for developing effectiveness assessment tools were focused primarily on measuring efficiency rather than effectiveness. Agency evaluation meant an assessment of agency's operations, its managerial performance and input measures as opposed to "outcome" measures. In UWASIS II (United Way of America, 1976), each of the 587 program definitions identified, suggested "program products." But in most cases the products were input or output measures: number of days of daycare provided, number of children adopted, number of hours of counseling provided, etc.

A number of local United Ways developed good manuals for agency evaluation. Several major national organizations also distributed such manuals to their affiliates. Some of these manuals were called "Self Assessment Guides." But again, these manuals were used for evaluating agency operations and not program effectiveness, which remained elusive to most. In the absence of practical, cost-effective program evaluation tools, greater reliance was placed on fiscal accountability and managerial effectiveness.

Notwithstanding the above, there seems to be a resurgence of interest in program evaluation, locally, nationally and even internationally. Increasing competition for tax as well as contributed dollars and scarce resources prompt donors and funders to ask once again: What good did the donation produce? What difference did a foundation grant or United Way allocation make in the lives of those affected by the service funded?

Particularly, in the political arena, there seems to be a great sense of urgency and frustration with regard to finding some indication of success (or lack thereof) of a whole variety of social programs aimed at helping individuals and families for which government spends billions of dollars annually. Taxpayers have a right to know whether their taxes are helping to improve the condition of the neediest amongst us.

II. SURVEY METHODOLOGY

In November of 1990 surveys regarding activities in evaluation were sent to 186 organizations. The sample was a purposive sample drawn by United Way of America staff. Organizations in the sample were intentionally selected for their representation of a larger group of non-profit organizations, either as a funder or as the national representative of a group of organizations (such as Girls Scouts of the U.S.A.).

The sample consisted of the following organizations:

- 25 largest foundations according to grant amount,
- 25 largest community foundations, according to grant amount,
- 25 largest United Ways, according to amount raised,
- 37 largest national social service agencies,
- 20 largest national health agencies,
- an additional 54 national organizations with membership in the INDEPENDENT SECTOR, representing educational, environmental, health, social service and arts and cultural organizations.

Responses were received from 91 organizations for a response rate of 48.9%. Respondents were divided among types of organizations as follows:

Foundations	26%
United Ways	22%
Social Service	19%
Health	16%
Education	9%
Other	9%

Respondents from arts/culture environmental organizations were coded as "Other" due to the low number of respondents in these groups alone.

III. CONCLUSIONS

- There was consensus across respondents that evaluation is beneficial and necessary. This was specifically true in the case of evaluation of Program Outcomes/Results.

- The focus of evaluation activity in the non-profit sector is on measurement of Volume of Program Delivery and Compliance with Standards. The lowest areas of activity are assessment of Program Outcomes/Results and Participant Satisfaction.

- The focus of assistance provided by funders and national organizations is on Management Assessment and Program Outcomes/Results. It is interesting that these are NOT the most commonly carried out activities, indicating that our assistance is directed toward growth and change rather than status quo. It is also the case that these two areas are arguably the most difficult and require some expertise in management and evaluation practices, which non-profit organizations may be less likely to have internally.

- The types of assistance most widely available now are consultation by funder or national organization staff, and training/workshops.

- The two greatest barriers to conducting assessment of Program Outcomes/Results are perceived to be lack of funding and lack of skills among staff of non-profit organizations.

- The two most useful strategies to increase assessment of Program Outcomes/Results are increased willingness of funders to support it and training for staff. Access to an organization with expertise and better measures and tools were close to training as choices three and four.

- Some identifiable differences exist across different types of non-profit organizations.
  - Foundations are less likely to see funding as a barrier to conducting assessment of Program Outcomes/Results. They are not more likely to report this type of evaluation occurring among grantees, however. They identified the greatest barrier to this as lack of staff skills.
  - Foundations and United Ways report much fewer staff in evaluation than national social service and health organizations. This is probably due to the fact that the funders rely on staff of these types of organizations to conduct evaluation and additionally, that foundations rely on outside consultants for larger evaluations.
  - United Ways tend to report assessment of Volume of Program Delivery and Participant Characteristics more than other types of organizations, and to report less assessment of Program Outcomes/Results.
  - National Social Service organizations report the greatest amount of assistance available to members. They provide more assistance than others in assessing Management Practices, Compliance with Standards, Volume of Program Delivery and Participant Satisfaction.
  - Health organizations are much more likely to identify research to create better measures and instruments as a strategy to increase assessment of "Program Outcomes/Results," ranking it second to funding. They are much less likely to view additional training as a useful strategy.
  - The focus of national educational organizations seems to be largely on "Compliance with Standards." They report this occurring more often than other organization types, and this is the only area in which they are not less likely than all others to provide assistance.

- Those who have developed a systematic process of conducting assessment of Program Outcomes/Results have gone through a staged process, bringing grantees/members through the following evolution:
  - Appreciation
  - Understanding
  - Ability to Define Objectives
  - Ability to Design Evaluation Measures
  - Data Collection
  - Discussion of Results.

Each of these stages was achieved as a team, with consultation and training available at each step.

IV. RECOMMENDATIONS

- If funders wish for evaluation to occur, particularly the assessment of program outcomes/results, they must be willing to allow grantees to use funds for this purpose. This is especially the case since funders evidently do not have dedicated staff in their own organizations to conduct evaluation. Funding of evaluation should be viewed as a necessary investment in program improvement.

- Similarly, provider organizations and their boards must see assessment of outcomes/results as necessary and integral to their business, not as fluff to be added with "extra" money.

- A new paradigm of program evaluation, separate from that of rigorous evaluation research models, needs to be developed for non-profit organizations. The consistent focus on cost of evaluation among respondents demonstrates that outcome evaluation is regarded as something complex and costly. While we do not suggest that evaluation does not require dedication of financial and staff resources, it cannot be seen as something so prohibitively expensive and sophisticated that it is beyond reach. There are many quasi-experimental and qualitative evaluation methods that can be routinely and realistically applied in non-profit agencies. The development of these models should be furthered, rather than continuing a focus on "pure" evaluation research.

- Additional training in program outcome evaluation is needed for staff of funders and providers. This training is not desired in an academic setting, again reinforcing the notion of a new paradigm of applicable evaluation models. Additional research to develop sound measures and instruments, and access to an outside organization with expertise in the area, will compliment and increase application of skills taught in training workshops.

- Non-profit organizations must work in concert to develop training and resources to support program outcome evaluation. Currently, there is duplicate effort occurring as various national groups independently develop their own materials. There does not appear to be any clear national resource or training program that is meeting the need. A forum should be created where existing materials can be shared and opportunities for collaborative efforts identified and implemented.

- Funders and national organizations should work together to design outcome measures and instruments to be used by member organizations nationwide so that the funder expectations and available techniques will be compatible.

- Funders must work with their grantees to develop program evaluation strategies that make sense and can be implemented by providers. The approach to this must be implemented incrementally with training and mutual participation at each stage.

- More attention should be given to the assessment of participant satisfaction. This provides a very basic measure of the recipient's perception of the quality and benefit of services, and can generally be collected using simple, inexpensive, easy to interpret methods. Until program outcome data becomes more widely available, this type of evaluation can serve as a reasonable proxy.

BIBLIOGRAPHY

Evaluation of Youth Development Programs.

Big Brothers/Big Sisters of America. (no date). Agency self-assessment and evaluation. Philadelphia: Author.

Boys & Girls Clubs of America. (1991). Commitment to quality: A process for evaluating and improving core program. New York: Author.

Boys & Girls Clubs of America. (1991). The effects of Boys & Girls Clubs on alcohol and other drug use and related problems in public housing (Final Research Report). New York: Author.

Case, R. (1987). How are we doing? A framework for evaluating development education programs. Washington, D.C.: InterAction.

Cohen, S., & Wilson-Brewer, R. (1991, September). Violence prevention for young adolescents: The state of the art of program evaluation (Working Paper). Washington, D.C.: Carnegie Council on Adolescent Development.

Cook, T. A., & Anson, A. R. (in press). From causal description to causal explanation: Three exemplary evaluations of adolescent health programs. In S. G. Millstein, E. O. Nightingale, & A. C. Petersen (Eds.), Adolescent Health Promotion. Cambridge, MA: Oxford University Press.

Ellickson, P. L., & Bell, R. M. (1990, March 16). Drug prevention in junior high: A multi-site longitudinal test. Science, 247, 1299-1305

Evaluation and Training Institute. (1988). DARE longitudinal evaluation report. Los Angeles, CA: Author.

Girls Incorporated. (1991). Assess for success: Needs assessment and evaluation guide. New York: Author.

Girl Scouts of the U.S.A. (1991). Girl Scout Council self-evaluation. New York: Author.

McLaughlin, M. W. (1987). Evaluating social programs: What's been learned. Unpublished manuscript, Stanford University.

The Milton S. Eisenhower Foundation. (1991). Issues in evaluation for discussion at the second international conference on urban safety, drugs and crime prevention. Washington, D.C.: Author.

Moberg, D. P. (1984). *Evaluation of prevention programs: A basic guide for practitioners*. Madison: Wisconsin Clearinghouse.

Project Dare: How's it Doing? (1990, Spring). *Prevention File*, p. 16-18.

Schorr, L. B. (1988). *Within our reach: Breaking the cycle of disadvantage*. New York: Doubleday.

Smith, C., & Kennedy, S. D. (1991). *Final impact evaluation of The Friendly PEERsusasion Program of Girls Incorporated*. New York: Girls Incorporated.

Stake, R. E. (1986). *Quieting reform*. Chicago: University of Illinois Press.

Sumariwalla, R. D., & Taylor, M. E. (1991, March). *The application of program evaluation in the management of the non-profit sector: An exploratory study*. Paper prepared for Spring Research Forum, "Leadership and Management," Cleveland, Ohio.

Torney-Purta, J. (1987). *Improving education through evaluation: Assessing global education programs and outcomes*. (Available from Judith Torney-Purta, Institute for Child Study, University of Maryland, College Park, MD 20742).

Wilson-Brewer, R., Cohen, S., O'Donnell, L., & Goodman, I. (1991, September). *Violence prevention for young adolescents: A survey of the state of the art*, (Working Paper). Washington, D.C.: Carnegie Council on Adolescent Development.

Wood, M. M. (1991, March). *A theory-driven framework for evaluating youth programs* (ISPS Working Paper No. 2163). New Haven, CT: Yale University, Institution for Social and Policy Studies.

YWCA of the U.S.A. (1989). *Taking the lead with PACT: Peer education in sexuality and health*. New York: Author.